

AD-R207-577

MULTIPLE SENSOR FUSION FOR DETECTING TARGETS IN FLIR

172

(FORWARD-LOOKING INF. (U) AIR FORCE INST OF TECH

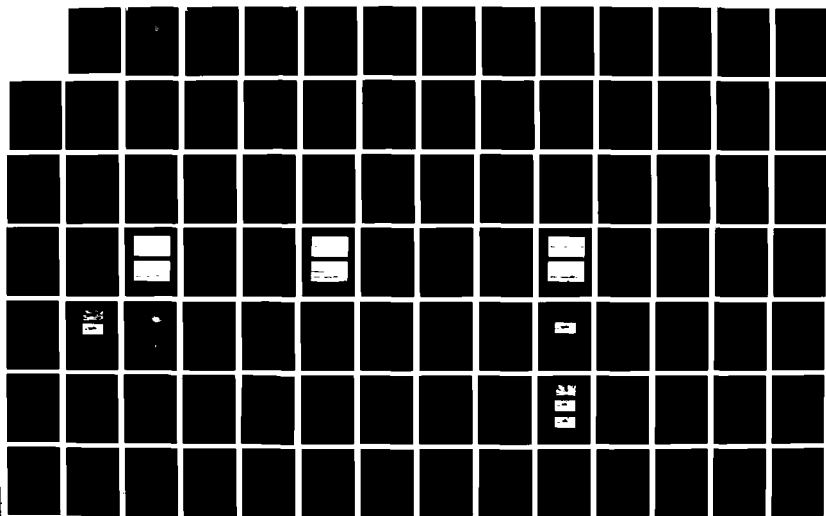
WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI. M C ROGGENMANN

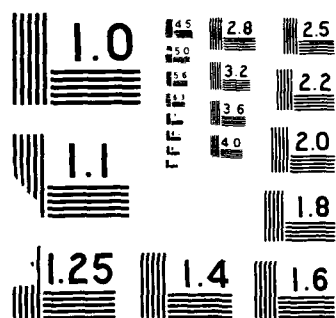
UNCLASSIFIED

MAY 89 AFIT/DS/ENP/89-1

F/G 17/5.1

NL





AFIT/DS/ENP/89-1

1

AD-A207 577

DTIC  
ELECTE  
MAY 05 1989  
S D & D

MULTIPLE SENSOR FUSION FOR DETECTING  
TARGETS IN FLIR AND RANGE IMAGES

DISSERTATION

Michael C. Roggemann

Captain, USAF

AFIT/DS/ENP/89-1

Approved for public release; distribution unlimited.

89 05 050

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

|   |       |  |  |   |                                |
|---|-------|--|--|---|--------------------------------|
| 1a. REPORT SECURITY CLASSIFICATION<br>UNCLASSIFIED  |       |  | 1b. RESTRICTIVE MARKINGS<br>NONE   |   |                                |
| 2a. SECURITY CLASSIFICATION AUTHORITY<br>N/A  |       |  | 3. DISTRIBUTION / AVAILABILITY OF REPORT<br>Approved for public release; distribution unlimited.                     |   |                                |
| 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE<br>N/A  |       |  |  |   |                                |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S)<br>AFIT/DS/ENP/89-1   |       |  | 5. MONITORING ORGANIZATION REPORT NUMBER(S)  |   |                                |
| 6a. NAME OF PERFORMING ORGANIZATION<br>School of Engineering  |       | 6b. OFFICE SYMBOL<br>(If applicable)<br>AFIT/ENP |  | 7a. NAME OF MONITORING ORGANIZATION               |                                |
| 6c. ADDRESS (City, State, and ZIP Code)<br>Air Force Institute of Technology<br>Wright-Patterson AFB OH 45433-6583  |       |  | 7b. ADDRESS (City, State, and ZIP Code)  |   |                                |
| 8a. NAME OF FUNDING / SPONSORING ORGANIZATION   |       | 8b. OFFICE SYMBOL<br>(If applicable)             |  | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER   |                                |
| 8c. ADDRESS (City, State, and ZIP Code)   |       |  | 10. SOURCE OF FUNDING NUMBERS  |   |                                |
|   |       |  | PROGRAM<br>ELEMENT NO.   | PROJECT<br>NO.                                    | TASK<br>NO.                    |
|   |       |  | WORK UNIT<br>ACCESSION NO.   |   |                                |
| 11. TITLE (Include Security Classification)<br>See Box 19   |       |  |  |   |                                |
| 12. PERSONAL AUTHOR(S)<br>Michael C. Roggemann, M.S.E.E., Capt., USAF   |       |  |  |   |                                |
| 13a. TYPE OF REPORT<br>Ph.D. Dissertation   |       | 13b. TIME COVERED<br>FROM _____ TO _____         |  | 14. DATE OF REPORT (Year, Month, Day)<br>1989 May |                                |
| 15. PAGE COUNT<br>129   |       |  |  |   |                                |
| 16. SUPPLEMENTARY NOTATION  |       |  |  |   |                                |
| 17. COSATI CODES  |       |  | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)                                    |   |                                |
| FIELD   | GROUP | SUB-GROUP  |  |   |                                |
| 12  | 09    |  | Multiple sensor fusion, FLIR image processing, Range image processing, Bayesian decision theory, Pattern recognition |   |                                |
| 19. ABSTRACT (Continue on reverse if necessary and identify by block number)  |       |  |  |   |                                |
| TITLE: MULTIPLE SENSOR FUSION FOR DETECTING TARGETS<br>IN FLIR AND RANGE IMAGES   |       |  |  |   |                                |
| Dissertation Chairman: James P. Mills, Lt Col, USAF<br>Adjunct Associate Professor of Engineering Physics   |       |  |  |   |                                |
| 20. DISTRIBUTION / AVAILABILITY OF ABSTRACT<br><input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS |       |  | 21. ABSTRACT SECURITY CLASSIFICATION<br>UNCLASSIFIED   |   |                                |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>James P. Mills, Lt Col, USAF   |       |  | 22b. TELEPHONE (Include Area Code)<br>(513)-255-2960   |   | 22c. OFFICE SYMBOL<br>ASD/ENAM |

(continuation of block 19)

Automatic detection of tactical targets in corresponding sets of non-pixel registered forward-looking infrared (FLIR) sensor images and range sensor images was studied. A processing architecture was developed to address the problems associated with processing non-pixel registered imagery. The architecture used specialized sensor-dependent processing to segment the images, measure features, and analyze the single sensor feature data. The multiple sensor processes of geometric registration, multiple sensor feature measurement, and multiple sensor target detection were then applied.

Sensor-dependent segmentation processes passed a large fraction of the targets present in the imagery, along with a larger number of regions which did not correspond to any target. FLIR images were segmented based on pixel brightness. A new range image segmentation algorithm was developed which exploited the small-scale planarity of tactical vehicles. The post-segmentation target detection problem was that of partitioning segmented targets from segmented non-target regions. Feature information was processed to accomplish this task. The Bayesian minimum error criterion was adopted as the decision rule.

Two single sensor detection algorithms (FLIR-only and range-only) and three multiple sensor detection algorithms (FLIR assisted by range, FLIR/range; range assisted by FLIR, range/FLIR; and a single decision algorithm) were implemented. A novel multiple sensor feature, called the correspondence feature, was developed to exploit the observation that targets occupy the same space in all sensor views of a scene, while segmented non-target regions behave in this manner much less frequently.

➤ When performance was optimized for all cases, the multiple sensor approaches were found to provide improved performance in all comparative performance measures. In addition, the single decision algorithm was shown to detect more targets than any of the other detection algorithms. These results support the hypothesis that use of multiple sensors in future targeting systems will be advantageous.

MULTIPLE SENSOR FUSION FOR DETECTING  
TARGETS IN FLIR AND RANGE IMAGES

DISSERTATION

Presented to the Faculty of the School of Engineering  
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the  
Requirements for the Degree of  
Doctor of Philosophy

Michael C. Roggemann, B.S., M.S.

Captain, USAF

May 1989

|                    |  |
|--------------------|--|
| Accession For      |  |
| NTIS CRA&I         | <input checked="checked" type="checkbox"/> |
| DTIC TAB           | <input type="checkbox"/>                   |
| Unannounced        | <input type="checkbox"/>                   |
| Justification      |  |
| By                 |  |
| Distribution /     |  |
| Availability Codes |  |
| Dist               | Avail and/or<br>Special                    |
| A-1                |  |

Approved for public release; distribution unlimited.

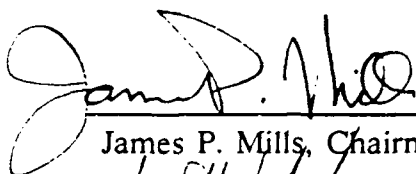


MULTIPLE SENSOR FUSION FOR DETECTING  
TARGETS IN FLIR AND RANGE IMAGES

Michael C. Roggemann, B.S., M.S.

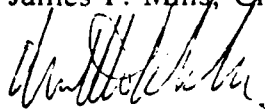
Captain, USAF

Approved:



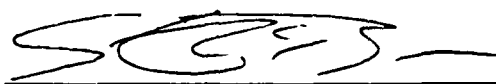
James P. Mills, Chairman

27 Apr 89



Matthew Kabrisky

27 Apr 89



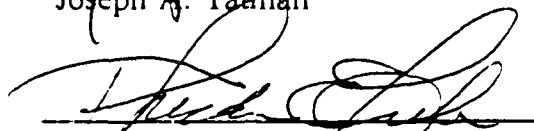
Steven K. Rogers

27 Apr 1989



Joseph A. Tatman

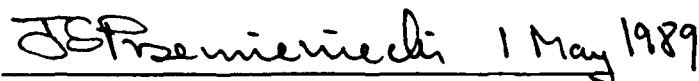
27 Apr 89



Theodore E. Luke

27 Apr 89

Accepted:



J.S. Przemieniecki  
Dean, School of Engineering

## ACKNOWLEDGEMENTS

In performing this research and writing this dissertation I received a great deal of help from many people. I am deeply indebted to my research committee chairman and faculty advisor, Lt Col James P. Mills, for his patience and steady guidance during all phases of my work. I also wish to express my gratitude to the other members of my research committee: Dr. Matthew Kabrisky, Capt Steven K. Rogers, and Capt Joseph A. Tatman. Throughout this project I found the collective guidance of my research committee to be consistently insightful and helpful. The data for this experiment was provided by the Army Center for Night Vision and Electro-Optics (CNVEO), Ft. Belvoir, VA. Mr. Joseph Swistak and Mr. John Nettleton, both of CNVEO, were instrumental in providing the data, and insights into sensor performance and data collection methods. Mr. Dan Zambon, Systems Engineer for the AFIT Information Sciences Laboratory, also provided a great deal of assistance in the area of computer support. Finally, I wish to thank my wife, Cynthia, and my son, Kent, for their loving understanding and support through all phases of my research.

Michael C. Roggemann



## Table of Contents

|   |      |
|---|------|
| Dissertation Approval Page .....                                    | ii   |
| Acknowledgements .....  | iii  |
| List of Figures .....   | vi   |
| List of Tables .....  | viii |
| Abstract .....  | ix   |
| I. Introduction .....   | 1    |
| 1.0 Problem Statement .....   | 1    |
| 1.1 Motivation .....  | 4    |
| 1.2 Approach .....  | 5    |
| 1.3 Background .....  | 10   |
| 1.4 Significant Results .....                                       | 12   |
| 1.5 Organization of the Dissertation .....                          | 17   |
| II. Architecture for Fusing Information from Multiple Sensors ..... | 18   |
| 2.0 Introduction .....  | 18   |
| 2.1 Architecture for Processing Multiple Sensor Information .....   | 21   |
| 2.2 Conclusions .....   | 25   |
| III. FLIR Image Segmentation .....                                  | 27   |
| 3.0 Introduction .....  | 27   |
| 3.1 Background .....  | 28   |
| 3.2 Segmentation Algorithm .....                                    | 29   |
| 3.3 Algorithm Performance and Scoring Method .....                  | 39   |
| 3.4 Limits of the Algorithm .....                                   | 40   |
| 3.5 Conclusions .....   | 40   |
| IV. Range Image Segmentation .....                                  | 42   |
| 4.0 Introduction .....  | 42   |
| 4.1 Background .....  | 46   |
| 4.2 Segmentation Algorithm .....                                    | 47   |
| 4.3 Error Threshold Selection .....                                 | 53   |
| 4.3.1 Derivation of $\mu_e$ and $\sigma_e$ .....                    | 55   |
| 4.3.2 Error Threshold .....   | 58   |
| 4.3.3 Physical Considerations for $\sigma_n(\rho)$ .....            | 61   |
| 4.4 Heuristics for Segmenting Range Images .....                    | 63   |
| 4.5 Algorithm Performance and Scoring .....                         | 65   |
| 4.6 Limits of the Algorithm .....                                   | 67   |
| 4.7 Conclusions .....   | 67   |
| V. Features and Geometric Registration .....                        | 69   |
| 5.0 Introduction .....  | 69   |

|  |     |
|--|-----|
| 5.1 Background .....   | 71  |
| 5.2 Feature Selection Method .....                                   | 71  |
| 5.3 Geometric Registration .....                                     | 76  |
| 5.4 Correspondence Feature .....                                     | 77  |
| 5.5 Conclusions .....  | 82  |
| VI. Single Sensor And Multiple Sensor Target Detection .....         | 86  |
| 6.0 Introduction .....   | 86  |
| 6.1 Background and Approach .....                                    | 88  |
| 6.2 Formulation of the Bayesian Class Estimation Problem .....       | 91  |
| 6.3 Training, Testing, and Performance Measures .....                | 92  |
| 6.4 Detection Algorithm Implementation .....                         | 95  |
| 6.5 Image Truth and Data Base Considerations .....                   | 99  |
| 6.6 Optimum Features Sets and Performance .....                      | 101 |
| 6.7 Conclusions .....  | 105 |
| VII. Conclusions and Future Directions .....                         | 106 |
| 7.0 Conclusions .....  | 106 |
| 7.1 Future Directions .....  | 109 |
| Appendix A: Sensor Description and Data Collection Methodology ..... | 111 |
| A.0 Introduction .....   | 111 |
| A.1 FLIR Sensor .....  | 111 |
| A.2 FLIR Data Collection Methods .....                               | 112 |
| A.3 Range Sensor .....   | 112 |
| A.4 Multisensor Data Collection Technique .....                      | 113 |
| Appendix B: Multiple Sensor Data Files .....                         | 115 |
| B.0 Introduction .....   | 115 |
| B.1 Data Base .....  | 115 |
| B.2 Selection Criterion .....  | 115 |
| Appendix C: Absolute Performance of Detection Algorithms .....       | 121 |
| C.0 Introduction .....   | 121 |
| C.1 Performance .....  | 121 |
| Bibliography .....   | 125 |
| Vita .....   | 129 |

## List of Figures

|  |    |
|--|----|
| Figure   |    |
| (1-1). Proposed architecture for multiple sensor automatic target detection system ..... | 3  |
| (1-2). Summary of target detection and false alarm performance .....                     | 15 |
| (1-3). Geometric relationship between images .....                                       | 16 |
| (2-1). Detailed multiple sensor processing architecture .....                            | 19 |
| (3-1). Block diagram of FLIR image segmentation algorithm .....                          | 30 |
| (3-2). Typical images from the FLIR image data base .....                                | 31 |
| (3-3). Histogram of FLIR image .....   | 32 |
| (3-4). Post-threshold FLIR images .....  | 34 |
| (3-5). Window used in recovery of lost pixels .....                                      | 37 |
| (3-6). Final segmented version of FLIR images .....                                      | 38 |
| (4-1). Smoothed and segmented range image .....  | 44 |
| (4-2). Plane parameters $b$ vs. $a$ .....  | 45 |
| (4-3). Block diagram of range image segmentation algorithm .....                         | 49 |
| (4-4). Range image sensing geometry and coordinate system .....                          | 50 |
| (4-5). Silhouette of range image resulting from application of error threshold .....     | 52 |
| (4-6). Error threshold as a function of range .....                                      | 60 |
| (4-7). Standard deviation of range measurements as a function of range .....             | 63 |
| (4-8). Additional example of range image segmentation performance .....                  | 66 |
| (5-1). Class-conditioned PDFs for range image length-to-width ratio feature .....        | 74 |
| (5-2). Class-conditioned PDFs for FLIR image correspondence feature .....                | 82 |

|   |    |
|---|----|
| (5-3). Class-conditioned PDFs for range image<br>correspondence feature ..... | 83 |
|---|----|

## List of Tables

|   |     |
|---|-----|
| Table   |     |
| (4-1). Error statistics for plane parameters .....          | 54  |
| (4-2). Comparison of error thresholds .....                 | 61  |
| (5-1). Rank-ordered $P_e$ for FLIR<br>image features .....  | 75  |
| (5-2). Rank-ordered $P_e$ for range<br>image features ..... | 76  |
| (6-1). Features giving minimum $P_e(tot)$ .....             | 101 |
| (6-2). Features giving maximum $P_d$ .....                  | 102 |
| (6-3). Features giving minimum $FAR$ .....                  | 102 |
| (6-4). Performance achieved with minimum $P_e(tot)$ .....   | 102 |
| (6-5). Performance achieved with maximum $P_d$ .....        | 103 |
| (6-6). Performance achieved with minimum $FAR$ .....        | 104 |
| (A-1). Range sensor image parameters .....                  | 113 |
| (B-1). DF1971 multiple sensor data sets .....               | 117 |
| (B-2). DF1671 multiple sensor data sets .....               | 118 |
| (B-3). DF1572 multiple sensor data sets .....               | 119 |
| (B-4). DF1771 multiple sensor data sets .....               | 120 |
| (C-1). Performance for FLIR:1; and Range:1 .....            | 122 |
| (C-2). Performance for FLIR:2; and Range:2 .....            | 122 |
| (C-3). Performance for FLIR:3; and Range:3 .....            | 122 |
| (C-4). Performance for FLIR:1,2; and Range:1,2 .....        | 123 |
| (C-5). Performance for FLIR:1,3; and Range:1,3 .....        | 123 |
| (C-6). Performance for FLIR:2,3; and Range:2,3 .....        | 123 |
| (C-7). Performance for FLIR:1,2,3; and Range:1,2,3 .....    | 124 |
| (C-8). Performance for FLIR:1,3; and Range:1 .....          | 124 |

## Abstract

Automatic detection of tactical targets in corresponding sets of non-pixel registered forward-looking infrared (FLIR) sensor images and range sensor images was studied. A processing architecture was developed to address the problems associated with processing non-pixel registered imagery. The architecture used specialized sensor-dependent processing to segment the images, measure features, and analyze the single sensor feature data. The multiple sensor processes of geometric registration, multiple sensor feature measurement, and multiple sensor target detection were then applied. Segmented regions were registered between the images, rather than pixels.

Sensor-dependent segmentation processes passed a large fraction of the targets present in the imagery, along with a larger number of regions which did not correspond to any target. FLIR images were segmented based on pixel brightness. A new range image segmentation algorithm was developed which exploited the small-scale planarity of tactical vehicles. The post-segmentation target detection problem was that of partitioning segmented targets from segmented non-target regions. Feature information was processed to accomplish this task. The Bayesian minimum error criterion was adopted as the decision rule.

Two single sensor detection algorithms (FLIR-only and range-only) and three multiple sensor detection algorithms (FLIR assisted by range, FLIR/range; range assisted by FLIR, range/FLIR; and a single decision algorithm) were implemented. A novel multiple sensor feature, called the correspondence feature, was developed to exploit the observation that targets occupy the same space in all sensor views of a scene, while segmented non-target regions behave in this manner much less frequently. Multiple sensor target detection algorithms were distinguished from single sensor detection algorithms by the

addition of correspondence feature information to the decision processes for the multiple sensor cases. Three comparative performance measures were used: (1) minimum error rate; (2) maximum detection rate; and (3) minimum rate of false alarms per detection declaration.

When performance was optimized for all cases, the multiple sensor approaches were found to provide improved performance in all comparative performance measures. In addition, the single decision algorithm was shown to detect more targets than any of the other detection algorithms. These results support the hypothesis that use of multiple sensors in future targeting systems will be advantageous.

# **MULTIPLE SENSOR FUSION FOR DETECTING TARGETS IN FLIR AND RANGE IMAGES**

## **I. Introduction**

### **1.0 Problem Statement**

The problem addressed in this dissertation is automatic multiple sensor target detection. Approaches to multiple sensor target detection were sought which were capable of overcoming some of the limitations of single sensor techniques and improving target detection performance compared to single sensor approaches. Evaluation of multiple sensor approaches to target detection and comparison of these approaches to single sensor approaches was also addressed. A data base of real, corresponding forward-looking infrared (FLIR) and absolute range images was used to develop and test multiple sensor techniques.

The goals of this research were to develop a general architecture for the extraction and use of multiple sensor information, and to develop and demonstrate multiple sensor processing approaches to improving target detection and false alarm performance. An additional goal was to compare single sensor and multiple sensor performance under equivalent conditions. A meaningful demonstration of the power of multiple sensor information processing was desired to develop a deeper understanding of how to extract and process multiple sensor information, to provide a concrete example of a working multiple sensor target detection system, and to provide evidence of performance improvements resulting from using multiple sensors. Testing of single and multiple sensor detection



systems under equivalent conditions was important for meaningful comparisons between single and multiple sensor approaches.

The initial hypotheses were that the performance of automatic target detection systems could be improved through the use of multiple sensors, and that the processing architecture shown in Figure (1-1) provided a general and useful approach to processing multiple sensor information. Performance improvements were expected by virtue of the additional information available in a multiple sensor system. The architecture provided a functional partitioning of the subproblems which was logical and sufficiently general to apply to other multiple sensor processing problems.

In the architecture of Figure (1-1) sensor-dependent processing was performed to locate potential target-bearing regions, called regions of interest, in the sensor images. Images which were non-zero only where regions of interest had been found were output to the feature measurement stage and to the image memory. The image memory and the data buffer were used to hold useful images and data for easy access in subsequent processes. Feature measurement, the act of converting pixel information about the regions of interest into numerical information, was also conducted on a sensor-dependent basis. The feature values were passed onto the sensor-dependent analysis block and to the data buffer. Sensor-dependent analysis consisted of computing class-conditioned probabilities for the single-sensor features observed, which were required by the detection algorithms. The multiple sensor decision and control processes geometrically registered the regions of interest, measured a novel multiple sensor feature, and performed the multiple sensor target detection processes. Though the sensor-dependent processes are shown for only one sensor in Figure (1-1), and this architecture was demonstrated for two sensors in this research project, the architecture should generalize directly to the case of more than two sensors.

The original contributions of this research lie in the development of a multiple sensor processing philosophy, validation of the processing architecture, demonstrated perfor-

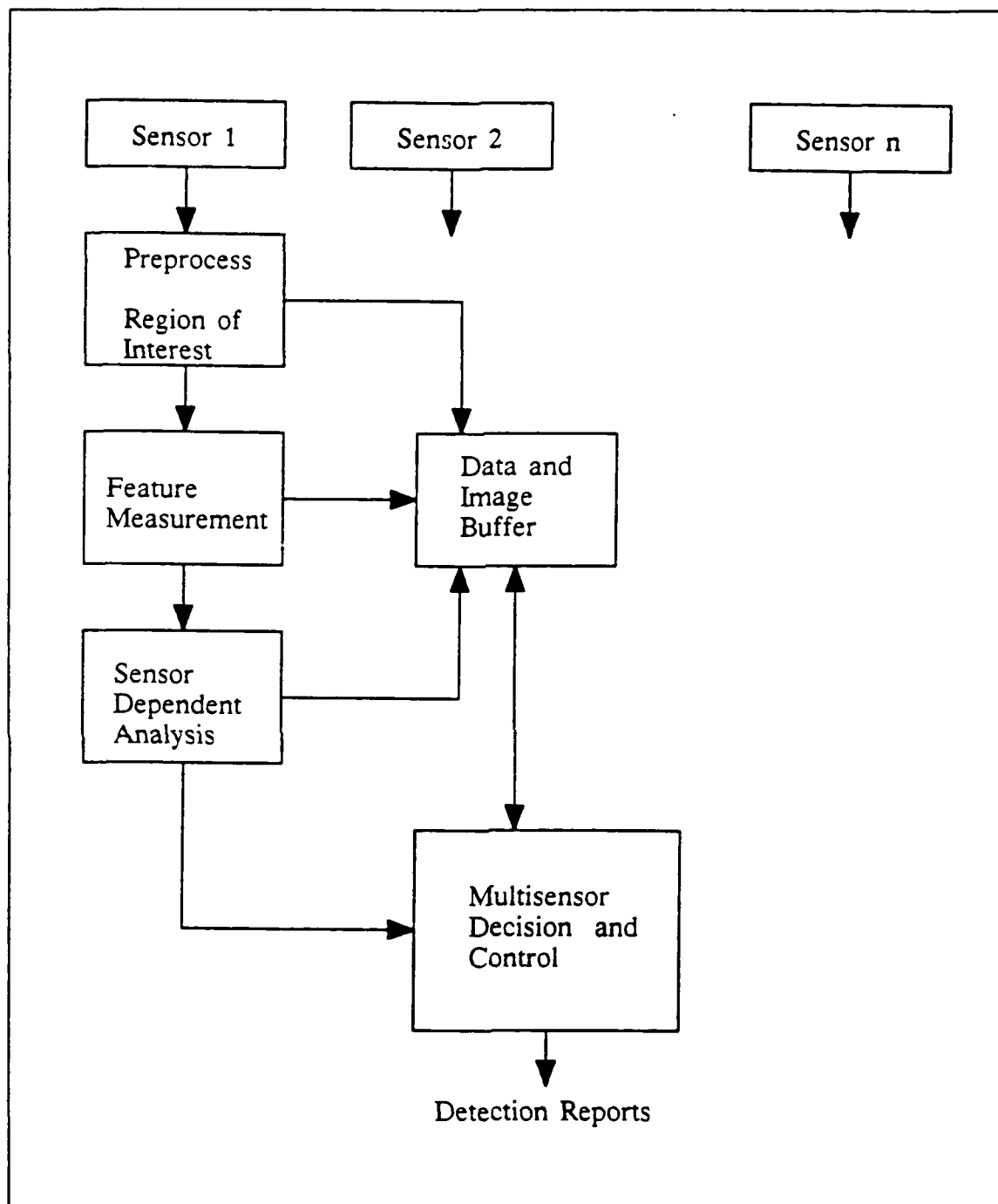


Figure (1-1). Proposed architecture for multiple sensor automatic target detection processing system.

mance improvements over single sensor approaches as a direct result of using a new multiple sensor feature, and certain aspects of the low-level processing of the sensor data.

Philosophical, theoretical, and implementation details of these topics are discussed in the succeeding chapters.

This work concerned the problem of automatically detecting man-made military vehicles in natural backgrounds only. No research was performed on the problem of recognizing the targets detected. However, an approach to multiple sensor target detection, such as the one developed here, would provide a useful target cue for input to a target recognition system. In a target recognition system, a target detection system would filter the scene for potential target-bearing regions which would be passed to the recognition system. The role of the recognition system would be to determine the class of segmented regions passed by the target detection system; for example, tank, truck, armored personnel carrier, or clutter. Multiple sensor information could also be used in the recognition process, but this work was not considered here.

Five sections remain in this introductory chapter. Motivating factors for this research project are discussed in the next section. This is followed by a summary of the approach. Background material pertinent to the general problem of multiple sensor information extraction and fusion is then presented. Next, the significant results of this research are summarized. The chapter concludes with an overview of the organization of the dissertation.

## **1.1 Motivation**

Research in the general area of automatic target detection is motivated by the desire to automate the process of detecting targets. Potential military applications of a viable automatic target detection technology include a wide range of manned fighting vehicles and unmanned missiles. Reliable automatic detection of targets is a step toward realizing targeting systems which require less, or no, human intervention.

It has been observed that current single sensor targeting technology is not capable of meeting projected operational requirements (Comparato, 1988). As a result, multiple

sensor approaches have been proposed to bolster targeting performance (Comparato, 1988; Duane, 1988; Roggemann et al, 1988; Ruck et al, 1988).

Multiple sensor systems should be capable of improved target detection performance, on average, when compared to single sensor systems by virtue of the additional information available to a multiple sensor detection system. Additionally, a multiple sensor system should provide some capability when one sensor is performing poorly due to imaging conditions, intentional countermeasures, or malfunctions, while a single sensor system would be severely limited or disabled under such conditions (Bullock et al, 1988; Comparato, 1988).

The scientific and engineering aspects of multiple sensor information extraction and fusion are active topics in the research community (Bullock et al, 1988; Comparato, 1988; Duane, 1988; Duda et al, 1979a; Magee and Aggarwal, 1985; Magee et al, 1985; Mitiche and Aggarwal, 1986; Roggemann et al 1988). Questions regarding the types of information to extract from multiple sensor data, registration of information in time and space, and information combination methods are unresolved for many applications. Thus, new approaches and results in nearly all aspects of multiple sensor information extraction and fusion are of interest to the research community.

## 1.2 Approach

The philosophy of performing multiple sensor processing developed and implemented here was that of partitioning sensor-dependent and multiple sensor processes. This approach was well suited to the case examined, where using multiple sensor information earlier in the target detection process would have been complicated by lack of pixel registration between the different sensor images. The individual strengths and the fundamentally different views of the scene provided by FLIR and range sensors were exploited by specialized low-level pixel processing and feature extraction algorithms. Geometric registration of the sensor-dependent information and measurement of multiple sensor information was performed after completion of the sensor-dependent processes,

using information derived from the sensor-dependent processes. Thus, the sensor dependent processes were concerned with the measurement of information about the scene based on their unique view of the scene. The multiple sensor processes were concerned with geometrically registering regions in the images, measuring information which could not be obtained from either sensor operating alone, and using this information to make decisions. This partitioning of functions would, in principle, allow a multiple sensor system to continue functioning in the presence of performance degrading conditions impacting one sensor.

The multiple sensor target detection problem was addressed in the following manner. A processing architecture suitable for processing multiple sensor imagery was developed. A data base of real, corresponding sets of high quality absolute range and FLIR data was obtained and used for algorithm development and testing. Segmentation algorithms for FLIR and range images, which extracted potential target-bearing regions in the imagery, were developed. The segmentation algorithms found most of the targets, but also passed a significant number of non-target regions. The post-segmentation target detection problem was modeled as a two class discrimination problem, with the classes being target and non-target. A set of features found to be suitable for the two class discrimination process was developed and computed for all segmented regions in the FLIR and range image data bases. A new multiple sensor feature called the correspondence feature, obtainable only through use of multiple sensors, was developed to provide additional information to the multiple sensor classification algorithms. The single and multiple sensor classification rules were implemented, tested, and compared. Each step in the approach is now summarized.

Modeling the target detection problem as a two class discrimination problem was a reasonable concession to the nature of image segmentation. The output of a segmentation system was an image composed entirely of zeroes except where potential target-bearing regions had been found. The segmentation algorithms were imperfect selectors of targets:

segmented images typically contained most of the targets which appeared in the image, but also passed a large number of non-target regions. The post-segmentation class discrimination problem was that of separating targets and non-targets through feature measurements and decision logic.

A multiple sensor processing architecture was developed which exploited the individual strengths of the sensors, and extracted and processed information available from multiple sensors. Extraction of single and multiple sensor information, geometric registration, two single sensor target detection techniques, and three multiple target detection techniques were implemented in the architecture. The processing system developed contained outputs for target detection and false alarm performance evaluation for FLIR-only, range-only, and three types of multiple sensor target detection algorithms. These outputs allowed easy comparisons between the various approaches.

The data base consisted of a set of 97 real FLIR images and 57 real range images. The data base was obtained from the Army Center for Night Vision and Electro-Optics, Ft. Belvoir, VA. It was collected as part of a larger effort to acquire a data base for the development and testing of automatic targeting systems for Army applications. Corresponding FLIR and range images in this data base were from colocated sensors: the images were not pixel-registered. The data provided was manually inspected to eliminate image sets which were unsatisfactory for sensor fusion research. Corresponding image sets were excluded from the sensor fusion data base due to high noise, most commonly in the range image, or the inability to choose a common reference point in both images for geometric registration purposes.

Sensor-dependent segmentation was accomplished through development and implementation of separate FLIR and range image segmentation algorithms. The tendency of sun warmed and exercised vehicles to appear brighter than the background was exploited by the FLIR segmentation algorithm. The observation that tactical military vehicles tend to be composed of small, approximately planar surfaces, while much of the background

does not possess this property, was exploited by the range image segmentation algorithm. Both the FLIR and the range image segmentation algorithms accurately segmented a large fraction of the targets observable in the imagery. Both segmentation algorithms also passed a number of non-target regions.

A system for consistently identifying segmented regions was developed, and the segmented images were manually inspected to obtain image truth. In the scheme for acquiring image truth, each segmented region was manually identified as either a target or a non-target, and this information was stored. The target detection systems were then tested by using the class estimation algorithm to obtain an estimate of the class membership of each segmented region and comparing the result to the image truth for that region. Image truth data was also used to compute class-conditioned probability density functions needed to train the classification algorithm.

An initial set of single sensor features were selected based on an evaluation of the individual sensor physics and the distinguishing characteristics of segmented target and non-target regions. The features used were insensitive to small changes in the pixels present in a segmented region; for example, the length-to-width ratio of a segmented region. Feature values were computed and stored for all segmented regions in the data base. A feature selection process was applied to select a subset of the initial feature set which provided optimal performance.

A novel multiple sensor feature, called the correspondence feature, was developed to add information to the multiple sensor class estimation process. The philosophy of the correspondence feature was that targets viewed by both sensors jointly occupy the same scene space, while segmented non-targets behave in this manner much less frequently.

Accurate geometric registration between the sensor images was required to measure the correspondence feature. A technique requiring a one-time manual review of the segmented images was developed to obtain the required registration. This technique allowed the selection of a single pixel in each of a corresponding set of FLIR and range

images, called the common pixel, to be selected which originated from approximately the same point in the scene. Location of corresponding positions between the images was then handled by computing angular displacements from the common pixel, a process called pixel translation. Common pixel locations for each pair of images were stored and accessed as needed.

Three multiple sensor detection techniques were developed: FLIR as the dominant sensor, called the 'FLIR looking into range' (FLIR/range) algorithm; range as the dominant sensor, called the 'range looking into FLIR' algorithm (range/FLIR); and an algorithm which provided a single decision for each parcel of space segmented by either sensor, called the 'single decision' (SD) algorithm. The FLIR/range and range/FLIR algorithms used the concept of the decision process in a dominant sensor image being assisted by information from the other sensor image, called the non-dominant sensor image. For example, in the FLIR/range algorithm the FLIR image was the dominant sensor image. The FLIR/range and range/FLIR algorithms were capable of declaring detections only on targets segmented in the dominant sensor image. Hence, if the dominant sensor failed to segment a target, that target was forever lost in these two approaches.

The SD algorithm did not possess this limitation, and was capable of correctly detecting targets which were segmented by only one sensor. The single decision algorithm required a technique for resolving cases where segmented regions in both images occupied the same region of space. A spatial deconfliction rule was developed to handle this problem.

A classic Bayesian approach was taken to the class estimation problem. The Bayesian minimum error decision criterion (Melsa and Cohn, 1978: 42; Devijver and Kittler, 1982: 33-43), called the Maximum a Posteriori (MAP) decision rule, was used. Class-conditioned probability density functions (PDF) computed for the features provided part of the information required to use the MAP approach. Prior densities for the classes, also required for the MAP approach, were assumed to be equally likely. The class-



current active issues in sensor fusion and the potential benefits of such work is available (Mitiche and Aggarwal, 1986). Common themes in the sensor fusion literature are the benefits derived from extracting and using additional information available from more than one sensor (Duda et al, 1979a; Garvey and Lowrance, 1981; Lowrance and Garvey, 1983; Haskins, 1984; Mitiche and Aggarwal, 1986; Bogler, 1987; Kreigman et al, 1987; Comparato, 1988; Roggemann et al, 1988; Ruck et al, 1988) and the ability to maintain some level of system performance in the presence of sensor failures or degradations (Comparato, 1988; Bullock et al. 1988). Performance improvements through extraction and use of additional information available in a multiple sensor environment was the principal focus of this research.

A critical aspect of implementing any multiple sensor processing system is the need to register information obtained from the individual sensors in some geometrical space (Haskins, 1984; Mitiche and Aggarwal, 1986; Comparato, 1988; Roggemann et al, 1988). Registration allows information obtained from the sensors to be combined for appropriate regions of space. Pixel-registered sensors are not required for this process, but knowledge of the geometrical transformation between the various sensing coordinates is required (Haskins, 1984; Mitiche and Aggarwal, 1986).

Biological and mechanical examples of the use of multiple sensor information exist. The pit viper family of snakes integrate infrared sensing and vision to determine the correct striking angle (Mitiche and Aggarwal, 1986: 381), and humans routinely integrate information from a combination of sensory inputs to analyze their environment. Mechanical examples of the use of multiple sensor information include use of range and visible imagery to extract planar regions from office scenes (Duda, et al, 1979a), integration of various electronic warfare sensors with intelligence information to understand a threat environment (Garvey and Lowrance, 1981; Lowrance and Garvey, 1983; Bogler, 1987), and integration of stereo vision, range sensing, and tactile contact sensors to develop a 'world model' used in guiding a robot (Kreigman, et al, 1987). In both the biological and

the mechanical cases the additional information obtained from using more than one sensor enhances the performance of the system.

The cases of multiple sensor fusion cited above constitute evidence indicating that carefully designed multiple sensor systems can improve the performance of a multiple sensor system over the performance obtained by using only one sensor. In the absence of general results for choosing the sensors, types of information to extract, and a rule for combining and interpreting the information from multiple sensors, the ingenuity of the designer is taxed for each new application. Results of previous researchers were viewed as concrete examples of successful sensor fusion, providing inspiration and confidence that careful design and execution of multiple sensor algorithms could improve target detection performance.

None of the literature reviewed offered a direct solution to some of the subproblems defined during this project. In particular, suitable segmentation algorithms for FLIR and range images were not available, no specific set of sensor-dependent features have been defined, geometric registration between non-pixel registered images was not addressed, and no multiple sensor features were found in the literature. However, background information was found which helped structure approaches to the subproblems addressed above, and provided insight into other subproblems addressed during this research. Due to the diversity of the topics covered, the pertinent background material is discussed in the appropriate chapters.

#### **1.4 Significant Results**

Significant, and in some cases novel techniques and results, were developed in the course of this research. Specifically, a philosophy for processing multiple sensor information was developed and a general architecture for implementing this philosophy was successfully demonstrated; an effective FLIR image segmentation algorithm and a novel segmentation algorithm for segmenting tactical targets in range images were developed and demonstrated; a novel multiple sensor feature, called the correspondence feature,

was developed and shown to provide a powerful piece of information to the multiple sensor target detection process; use of the correspondence feature in conjunction with other features was shown to provide superior performance over either FLIR or range sensor performance alone; and an algorithm which declared a single decision for each parcel of space segmented by either sensor, the SD algorithm, was shown to detect targets which were not segmented in the dominant sensor image.

The utility of the philosophy of partitioning single and multiple sensor functions in a multiple sensor processing system which does not use pixel-registered sensors was validated. Performance improvements in detection and false alarm rates which resulted from implementing this philosophy provide strong evidence that performance improvements could be obtained for future systems through multiple sensor processing.

The FLIR segmentation algorithm is significant in the sense that reliable, high quality segmentation of targets was obtained. The FLIR segmentation algorithm passed approximately 91% of the targets found in the data base. Non-target regions appearing in the segmented images, or false segmentations, occurred at a rate of 0.58 per segmented region. Normalized on a per square degree of image space basis, false segmentations occurred in the FLIR image data base at a rate of 0.18 per square degree.

The range image segmentation system was unique in that target and non-target pixels are initially partitioned based on a novel planarity test. Use of planarity without regard for the orientation of the planes is a new and effective method for segmenting man-made vehicles from natural backgrounds in range image. The planarity test used was distinct from previous tests, and the critical parameter in the algorithm was shown to be approximated well by a function of readily obtainable, physically significant range sensing parameters. The planarity test automatically adapted to imaging and sensor performance measures and range. The range image target segmentation rate was approximately 88%. False segmentations occurred at a rate of 0.69 per segmented region, or at a rate of 1.61 per square degree.

The correspondence feature, which may only be measured in a multiple sensor environment, was developed to add information to the multiple sensor class estimation and algorithm control processes. The idea embodied in the correspondence feature was that targets jointly occupy the same space regardless of the sensing mode, while segmented non-target regions do not tend to behave in this manner. The correspondence feature measurement provided information about the joint spatial occupancy of segmented regions which was used in the class estimation process for all the multiple sensor algorithms, and in the spatial deconfliction process in the SD algorithm. The correspondence feature was an excellent feature because the FLIR and range segmentation systems tended to have false segmentations on different types of scene elements.

Use of correspondence feature information distinguished the information available to the multiple sensor target detection processes from the information available to the single sensor processes. All multiple sensor approaches developed provided improved performance over the single sensor cases in all the performance measures used. This result illustrates the benefits of multiple sensor processing for automatic target detection.

The best performance obtained, with all the systems optimized for maximum detection rate, is shown in Figure (1-2). The performance of five target detection algorithms is summarized in Figure (1-2): FLIR and range denote single sensor approaches; and FLIR/range, range/FLIR, and single decision denote multiple sensor approaches. Figure (1-2) shows that use of multiple sensor information was found to improve both target detection rates and the rate of false alarms per detector declaration.

In Figure (1-2), the detection rates for all algorithms are reported normalized to the number of segmented target regions appearing in the appropriate sensor image data base which were viewed completely by both sensor images. All segmented range image regions in the range image data base were viewed completely by the corresponding FLIR images. However, the converse was not true. A subset of all segmented FLIR regions was viewed by the range image data base. This situation is illustrated in Figure (1-3).

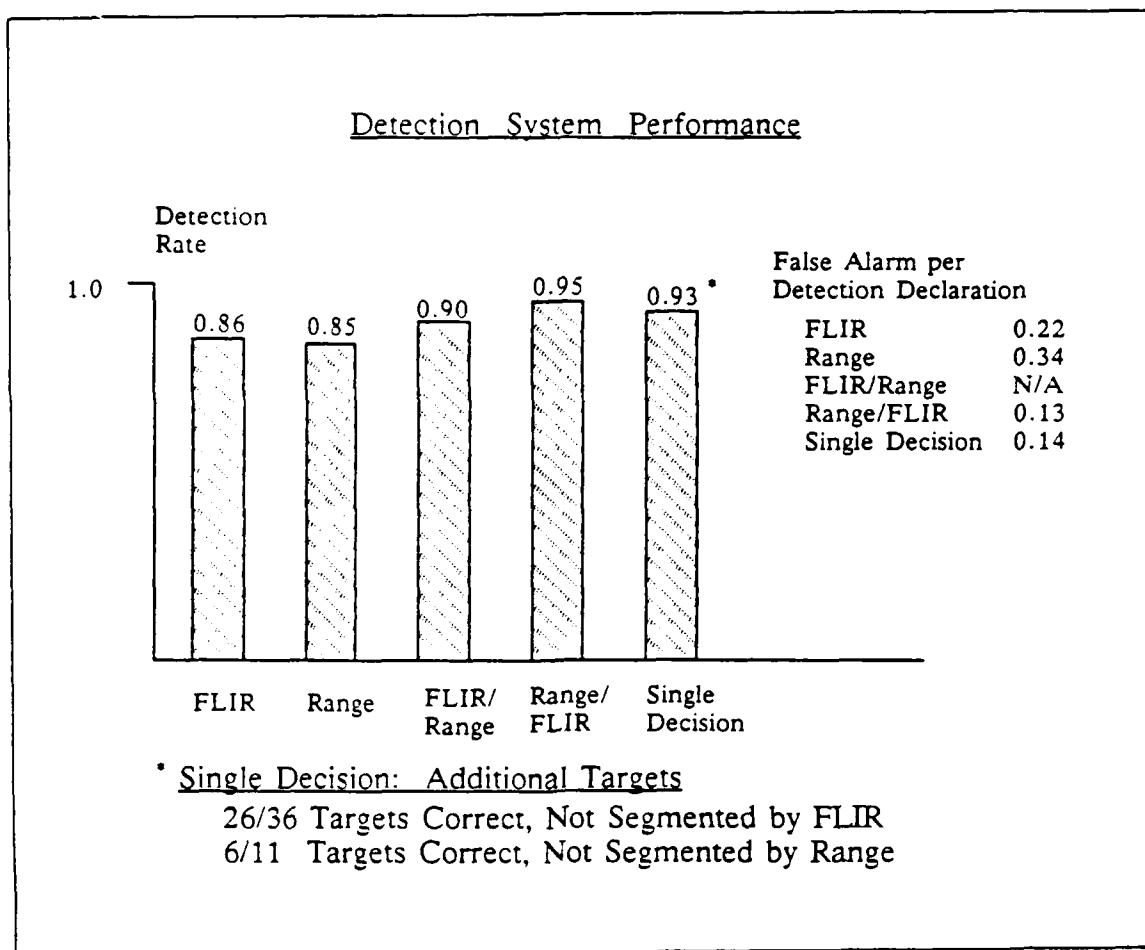


Figure (1-2). Summary of target detection and false alarm performance.

In Figure (1-3) the typical geometric relationship between FLIR and range image views of a scene is shown. The portion of the scene viewed by the range images was uniformly contained within the portion of the scene viewed by the FLIR images in the data base. Thus, there were more targets in the FLIR image data base than in the range image data base. However, only a subset of the FLIR image targets and potential false alarms were viewed completely by the range imagery. The results reported in this dissertation discuss only the subset of targets and potential false alarms viewed completely by both the FLIR and the range images.

Only a very small fraction of the potential FLIR image false alarms, 7.2%, was

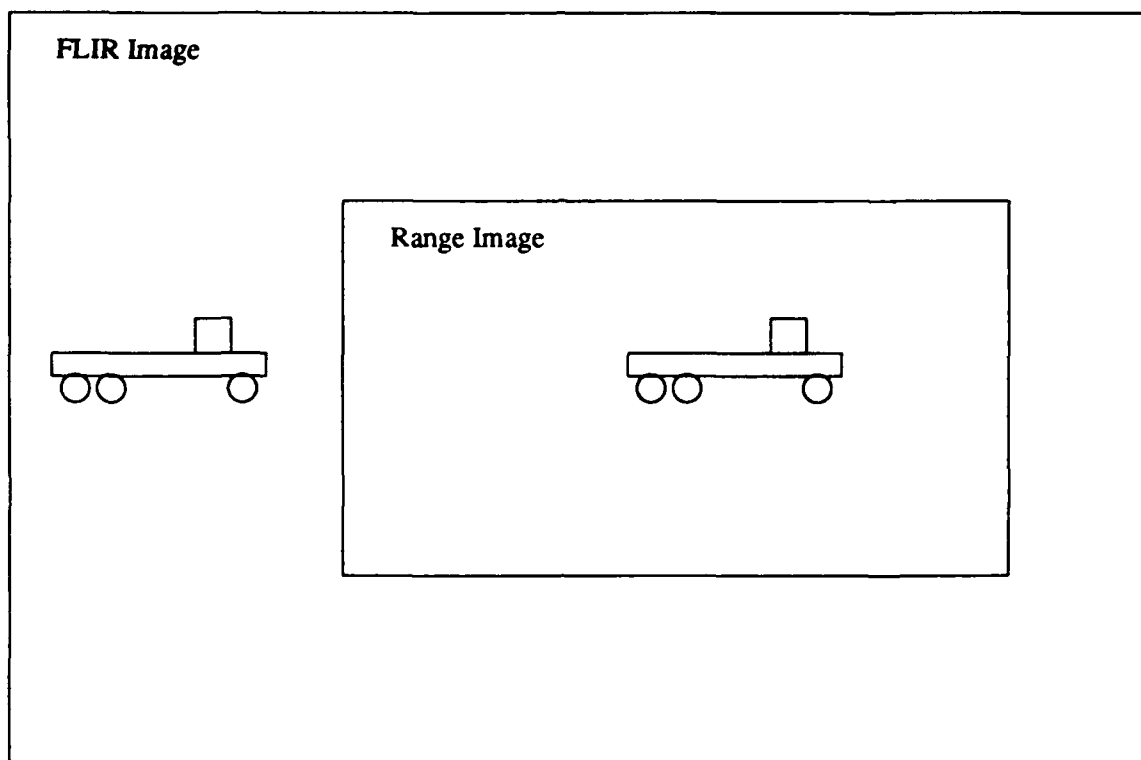


Figure (1-3). Geometric relationship between the portion of the scene viewed by the FLIR image and that viewed by the range image.

viewed completely by the range image data base. Thus, any estimate of the false alarm rate for FLIR images based on the set of potential FLIR image false alarms viewed completely by both sensors would be unduly low. Hence, the false alarm rate reported in Figure (1-2) is the result of FLIR-only performance on the entire data base of FLIR images. It is for this reason that no false alarm rate is reported for the FLIR/range algorithm in Figure (1-2).

Special note must be taken of the single decision algorithm performance. This algorithm was capable of detecting segmented target regions regardless of whether the regions appeared in both segmented images or in only one of the segmented images. The other detection approaches were limited to declaring target detections on segmented target regions which appeared in only one segmented sensor image. For example, the FLIR-only and FLIR/range algorithms were capable of detecting only the segmented

target regions which appeared in the segmented FLIR images. The single decision algorithm detected 26 of 36 target regions which were not segmented in the FLIR image data base and 6 of 11 target regions which were not segmented in the range image data base. A more detailed discussion of the data base and performance is contained in Chapter VI.

### **1.5 Organization of the Dissertation**

This dissertation is organized into six remaining chapters. Chapters II through VI contain technical discussions of the major problem areas addressed in the course of the research. Chapter VII provides conclusions and recommendations for future research.

Chapter II provides a discussion of the philosophy and implementation of the processing architecture used in this project. Chapters III and IV present the FLIR and range image segmentation algorithms, respectively. Chapter V discusses the selection and computation of single sensor features, geometric registration, and the multiple sensor correspondence feature. Chapter VI presents the class estimation decision rule, single and multiple sensor decision algorithms, and results computed under fair test conditions for all the detection algorithms developed.

## **II. Architecture for Fusing Information from Multiple Sensors**

### **2.0 Introduction**

The problem addressed in this chapter is that of defining a processing architecture for performing multiple sensor target detection. The main requirements for the processing architecture were that it be capable of extracting and merging information from corresponding FLIR and range images for the purpose of automatically detecting targets, and that it allow easy comparison of target detection approaches developed. This goal required the extraction of information from the sensor data, the preservation of useful information for later use, the ability to register information between different sensor views of the scene, the ability to gather additional information from one sensor image based on cues from another sensor, and the ability to perform a 'fair test' between competing approaches to target detection.

The processing architecture developed met these objectives. This architecture is shown in Figure (2-1). Individual strengths of the sensors were exploited in the sensor-dependent processes of segmentation, feature measurement, and sensor-dependent analysis. Useful information was retained in an image memory and a data buffer. A multiple sensor algorithm controlled the collection and use of multiple sensor information. Finally, the implementation allowed the computation and comparison of five different target detection schemes based on the sensors used in this research: FLIR only, range only, FLIR looking into range, range looking into FLIR, and the single decision case.

The philosophy of partitioning the sensor-dependent and multiple sensor processes was found to be well suited to the problem of processing non-pixel registered multiple sensor images. In this paradigm, the role of each of the sensor-dependent processing systems was to locate potential target-bearing regions through segmentation, and measure



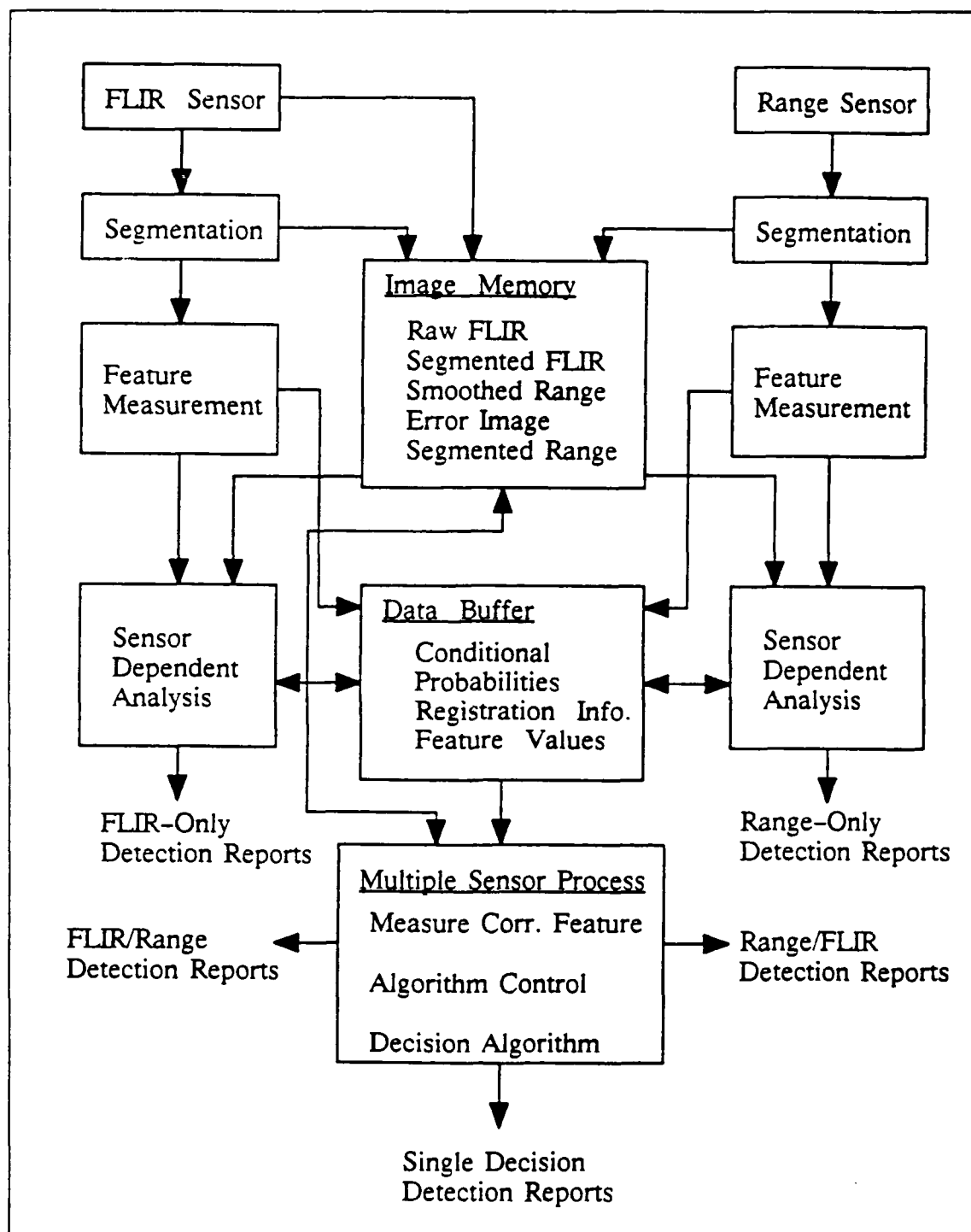


Figure (2-1). Detailed multiple sensor processing architecture.

features for these regions. Multiple sensor information was not processed until after specialized sensor-dependent processes had been applied to the raw sensor data. The role of the multiple sensor processing was to geometrically register the potential target-bearing regions found by the sensor dependent processes, measure multiple sensor information for these regions, and render a class estimate (target or non-target) for each region.

The case for partitioning sensor-dependent and multiple sensor processes follows from the lack of pixel registration between the images. Lack of pixel registration between the images would have made multiple sensor pixel level processes requiring precise registration (Duda et al, 1979a; Duane, 1988) very difficult, and none were attempted. The approach of registering segmented regions between the images, and searching these 'cued' regions, was adopted. Pixel level searches were conducted within cued regions to measure multiple sensor information. Allowances were made in the measurements for the possibility of small registration errors.

Though precise pixel registration between the sensors was not required in the multiple sensor processing system, a means of geometrically registering interesting regions, as determined by the sensor-dependent segmentation processes, was required. Geometrical registration of regions is a less stringent physical requirement on the sensors than the requirement of pixel registration between the sensors. Approaches using geometrical registration of non-pixel registered sensors allow each sensor used to be designed for optimal performance without the added physical constraints on the sensors necessary to obtain pixel registration. However, maintaining an accurate estimate of the geometrical transformation between the sensors would be required when multiple sensor operations are underway in non-pixel registered systems.

The computational burden associated with maintaining the geometric transformation between sensors in a non-pixel registered system is mitigated somewhat because multiple sensors which are not pixel registered may be used to search disjoint regions until multiple sensor information is required. Thus, the coverage of a non-pixel

registered multiple sensor system is, in principle, greater than that obtainable with an otherwise equivalent, but pixel registered multiple sensor system. The tradeoff between pixel registered and non-pixel registered multiple sensor systems must be made based on system performance requirements.

Two sections remain in this chapter. In the next section each functional block in Figure (2-1) is discussed in detail. The functional blocks are discussed in the context of information and data input/output. Implementation details are left to later chapters; the goal here is to explain the overall philosophy and functioning of the system. Conclusions are discussed in the final section of this chapter.

## **2.1 Architecture for Processing Multiple Sensor Information**

The processing architecture shown in Figure (2-1) is a refinement of the processing architecture originally proposed for this project, shown in Figure (1-1). This architecture provided a powerful approach to extracting and using information available from two sensors. The differences between the two figures resulted from knowledge gained in the course of the research. Though the implementation presented was developed for two sensors, the architecture should generalize directly to more than two sensors, as shown in Figure (1-1).

Six major functions are represented in the architecture: (1) sensing; (2) segmentation; (3) feature measurement; (4) memory; (5) geometrical registration and correspondence feature measurement; and (6) single and multiple sensor data analysis, control, and decision processes. Some of these processes were sensor-dependent in that they were performed using information available from only one sensor, while other processes used information obtained from both sensors. Sensor dependent functions were sensing, segmentation, single sensor feature measurement, and sensor-dependent data analysis. Multiple sensor functions were geometrical registration and correspondence feature measurement, multiple sensor process control, and multiple sensor class estimation.

Sensing was accomplished remotely from the processing. Descriptions of the sensors used and the data collection methods are provided in Appendix A. From an input/output perspective the sensors accepted their peculiar view of the scene as input and provided images as output. The images were two-dimensional arrays of numbers, where each entry in an array corresponded to the appropriate sensor's estimate of the sensed quantity for the scene element sampled. FLIR imagery provided estimates of the relative apparent temperature distribution in the scene. Range images provided estimates of the distance between the sensor and the scene element sampled. Sensor output was passed to the segmentation systems and to the image memory.

Differences in low level processes were required because FLIR and range images provide fundamentally different information about the scene observed. FLIR images provide a measure of the relative apparent temperature of each scene element (Lloyd, 1975:2-4), while range images provide a measure of the range from the sensor to each scene element (Bachman, 1979:79-120; Due and Peterson, 1982:215-226). Hence, the segmentation algorithms and, in some cases, the features measured for FLIR and range images were quite different.

Segmentation, a sensor-dependent process, had the goal of automatically extracting as many target regions as possible from the images while passing as few non-target regions as possible. The inputs to the segmentation processes were sensor images. Outputs consisted of segmented images, and in the case of the range image segmentation block, two useful intermediate images called the smoothed image and the error image. The computation and use of these intermediate images are discussed in Chapters IV and VI. Segmentation output was passed to the image memory and to the feature measurement block.

Segmented images were images composed entirely of zeroes except where regions passing all segmentation tests were found. The non-zero pixels in segmented images held the value of the corresponding pixel position in the raw or smoothed images for

FLIR and range images, respectively. The non-zero regions in segmented range images typically corresponded to the target pixels for most, or all, of the targets in the image, and some regions which did not correspond to any target. Thus, the post-segmentation target detection problem was reduced to partitioning the target regions from the non-target regions in segmented images.

Separation of targets and non-targets, called class estimation, was based on the measurement and analysis of feature information for each segmented region in both types of image. Two types of features were used to accomplish this task: single sensor features, and the multiple sensor correspondence feature.

Single sensor features were measured for each segmented region in both types of image. Input to the single sensor feature measurement processes consisted of the appropriate segmented image. In addition, the FLIR image feature measurement system required the raw FLIR image as input and the range image feature measurement system required the smoothed version of the range image as input. The features used were insensitive to small changes in the pixels present in segmented regions. An example of such a feature is the length-to-width ratio. Shape-related features were measured for both types of image. Also, brightness-related features were measured for FLIR images and distance-related features were measured for range images.

Output of the feature measurement process was an array of feature values indexed to a positive integer identifying each segmented region in both types of image. A system for consistently labeling the pixels in connected segmented regions was developed to make this approach feasible. These outputs were passed to the sensor-dependent analysis blocks and to a data buffer for later use.

The main function of sensor-dependent analysis of the feature information was to compute the class-conditioned probability of observing the combination of features measured for each segmented region. Feature values for each segmented region comprised the input to the sensor-dependent analysis blocks. Discrete class-conditioned

probability density functions (PDF), obtained from the training set and stored in the sensor-dependent analysis functional block, were used to obtain these class-conditioned probabilities. These probabilities were the most important output of the sensor-dependent analysis block. The outputs were passed to the data buffer for later use.

The sensor-dependent analysis function also made a single sensor estimate of the class (target or non-target) of each segmented region based on single sensor data, compared this estimate to image truth, and tabulated the results. The single-sensor decision criterion was identical to that used in the multiple sensor class estimation process. However, only information obtained from one sensor was used to make the single sensor class estimate.

The image memory held useful versions of the images as they were computed, avoiding the need to recompute them later. In the simulation environment available for this research the image memory consisted of memory arrays in a general purpose computer. Information retained in the image memory included the raw and segmented FLIR images, the segmented range image, and two useful intermediate images arising from the range image segmentation process, called the smoothed range image and the error image. The distinction is drawn between the image memory and the data buffer because the storage requirements for image memory are much larger than those of the data buffer.

The data buffer also held useful information which needed to be accessed subsequent to the process through which the information was obtained. Information in the data buffer consisted of the feature values and current estimates of the class conditioned probabilities for all segmented regions, and information required to geometrically register regions in the images.

The multiple sensor processes consisted of geometric registration of segmented regions between the images, measurement of the correspondence feature, obtaining information to resolve the joint spatial occupancy of segmented regions in both images in the single decision (SD) algorithm, and performing multiple sensor class estimation. Inputs

to the multiple sensor process block consisted of the contents of the image memory, the current estimate of the class-conditioned probabilities for each segmented region, and geometric registration information in the form of the common pixel. As output, this block provided the tabulated results of three multiple sensor decision algorithms: (1) the FLIR looking into range algorithm (FLIR/Range); (2) the range looking into FLIR algorithm (range/FLIR); and (3) the SD algorithm.

## 2.2 Conclusions

The architecture shown in Figure (2-1) provided a general and useful partitioning of functions in the multiple sensor target detection problem. Figure (2-1) is a detailed version of the proposed architecture shown in Figure (1-1). Refinements to Figure (1-1) shown in Figure (2-1) are the result of knowledge gained in the course of this project. This architecture allowed the implementation of the functional blocks to be addressed in relative isolation from the larger problem. Additionally, the architecture is, in principle, quite generally applicable. Variations on this architecture may appear in future systems.

The strength of this architecture derives from the separation of sensor-dependent processes and multiple sensor processes for non-pixel registered imagery. Sensor-dependent processing exploited information available from each sensor using algorithms developed specifically for that sensor. Multiple sensor processes were based on the outputs of the sensor-dependent processes, and multiple sensor information was obtained through the registration of interesting regions between the images.

In an operational system development the choice of whether to use pixel registered sensors or non-pixel registered sensors lies ultimately with the system designer. Pixel registered systems admit greater sophistication in the low level multiple sensor processes (for example, Duda et al, 1979a; Haskins, 1988) than can be accomplished using non-pixel registered sensors. Greater sophistication in these processes may contribute to improved performance, but this point has not been demonstrated through comparative studies. Non-pixel registered systems, in principle, would allow greater scene coverage

than an otherwise equivalent set of sensors which were pixel registered through use of a shared aperture. Greater scene coverage in a non-pixel registered system would result from careful design of the sensor scan patterns to cover different scenes until multiple sensor information was required to make a decision.



### **III. FLIR Image Segmentation**

#### **3.0 Introduction**

The problem addressed in this chapter is that of developing a technique to automatically find potential target-bearing regions in Forward-Looking Infrared (FLIR) sensor images. This process, called segmentation, had the goal of extracting the targets from the images as accurately and reliably as possible, while rejecting as much of the remainder of the images as possible.

The philosophy of FLIR image segmentation was to extract regions, or "blobs", whose edges and interiors closely corresponded to the visible bounds and interiors of the targets in the data base. Ideally, each target in a segmented image would have consisted of a region containing only target pixels, and no non-target regions would appear in the segmented images. The ideal case would have been quite difficult to achieve, and may be impossible to obtain. However, an algorithm which provided high quality segmentation was developed.

The algorithm developed was based on the following observations about the targets in the images: (1) the targets in the data base generally had a higher apparent temperature than the background, and thus appeared brighter than the background in the FLIR images; (2) the targets tended to be differentially heated due to operation and sun warming; and (3) the targets occupied a small fraction of the total pixels in the images. These observations indicated that target pixels and background pixels could be partitioned based on brightness, and suggested an adaptive method for performing this partitioning. The segmentation algorithm was based on a threshold operation, followed by a set of heuristic operations. The threshold was selected adaptively based on an automated inspection of the histogram of the median filtered version of the image. A set of heuristic operations were applied which were designed to reject additional non-target pixels

passed the threshold, and to recover a small number of target pixels which were inadvertently lost.

The algorithm was applied to a data base of 97 FLIR images found suitable for multiple sensor research, and scored. Optical parameters for the sensor, and data collection methods are discussed in Appendix A. The data sets and file names from the Army Center for Night Vision and Electro-Optics (CNVEO) June 1987 Multisensor Data Collection are listed in Appendix B. Appendix B also contains a discussion of criteria used to select FLIR and range image sets for the sensor fusion research.

Five sections remain in this chapter. Background pertinent to the segmentation algorithm is provided in the next section. This is followed by a discussion of the algorithm and its implementation. Performance of the algorithm on the data base and the scoring technique are then discussed. Limits to applying this algorithm are presented next. The final section contains conclusions drawn from the FLIR segmentation work.

### **3.1 Background**

FLIR images are two-dimensional arrays of numbers where each entry is a measure of the relative apparent temperature of the scene element sampled (Lloyd, 1975:1-4). Thus, to sense the presence of a target in FLIR imagery it is necessary that an observable apparent temperature difference exist between the targets and their immediate background (Lloyd, 1975:8). This criterion was largely met by the targets in the data base.

Partitioning of the pixels in an image into two classes based on brightness is often cast as the textbook problem of selecting a threshold (Gonzalez and Wintz, 1987:354-367). This approach was adopted here. In the present case, the two pixel classes were potential target and background.

When the brightness distributions of the two classes are known an optimal threshold may be selected (Gonzalez and Wintz, 1987:360-363). Unfortunately, no model for predicting these brightness distributions was known to exist. The problem was further

distinguished from the textbook case in that the brightness distributions for potential target and background regions were not separated by a poorly populated band of brightness levels.

A heuristic approach to selecting a threshold grounded in the observations about the brightness distributions of the targets and the background was adopted. The threshold selection algorithm and subsequent processing are discussed in the next section.

### 3.2 Segmentation Algorithm

The segmentation algorithm developed for FLIR images is shown in block diagram form in Figure (3-1). Raw FLIR images were initially median filtered to smooth spurious noise. The histogram of the median filtered image was then computed. A threshold computation algorithm was applied to the histogram. The threshold was applied to the median filtered image, creating an intermediate image called the post-threshold image. Heuristics were applied to the post-threshold image to remove unwanted pixels and recover a small number of pixels inadvertently lost during earlier stages of the algorithm.

Typical FLIR images from the data base are shown in Figure (3-2). Figure (3-2a) shows, from left to right, and M60A tank, an M113 armored personnel carrier (APC), a sandpaper covered target board placed for sensor calibration purposes, and a 2.5 ton truck. These targets were at a range of approximately 1070 m, and were viewed with the narrow field-of-view of the sensor. Figure (3-2b) is a wide field-of-view image which contains several targets: three clearly visible targets, two 2.5 ton trucks and one M60A tank at a range of 860 m; an M60A tank nearly matched in brightness to that of the background, lying to the left of the targets just mentioned, also at 860 m; and three targets appearing as an approximately equally-spaced array of three small bright spots, which were at a range of approximately 1700 m.

The histogram for the median filtered version of the image shown in Figure (3-2a) is shown in Figure (3-3). Figure (3-3a) shows the histogram on a scale sufficient to view

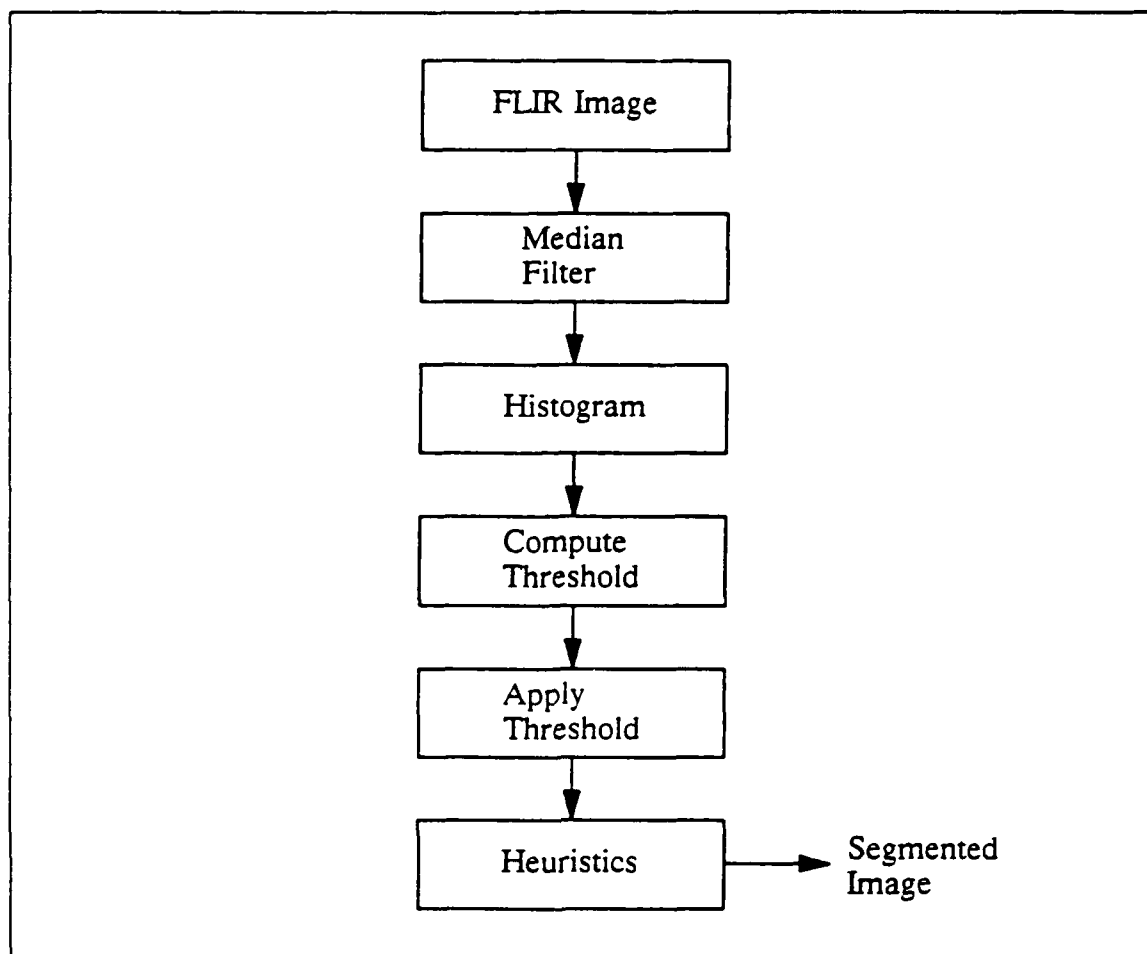
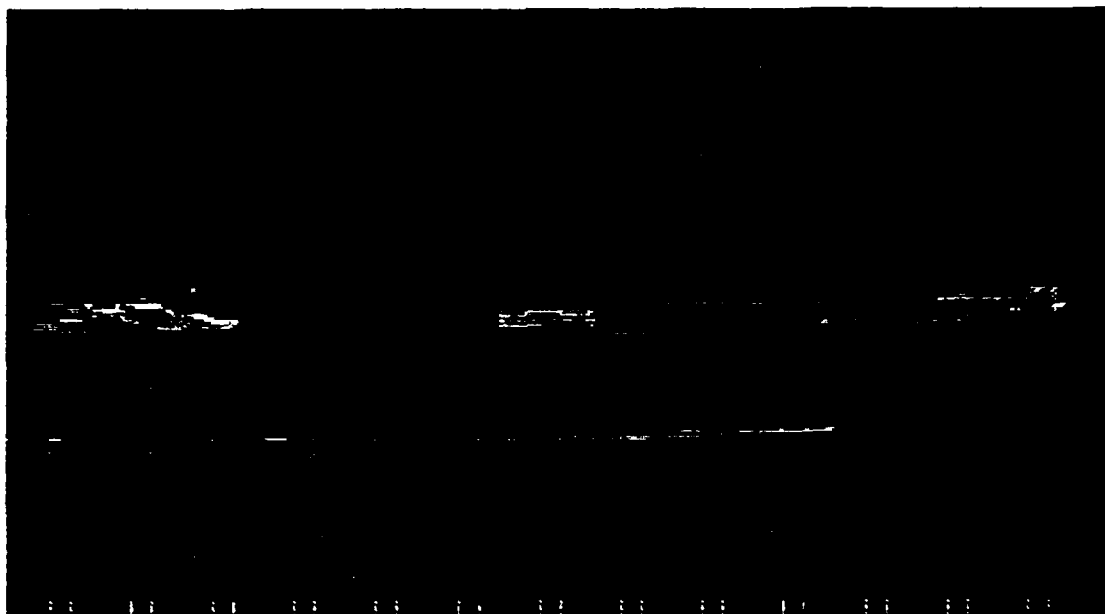


Figure (3-1). Block diagram of the FLIR image segmentation algorithm.

the entire histogram. Figure (3-3b) shows the same histogram, but with the vertical axis stretched to show the details of the relatively poorly populated levels.

The observations about the targets discussed in Section 3.0 contributed directly to the formulation of the segmentation algorithm. The observation that the targets were generally brighter than the background indicated that a threshold could be used as the initial step in segmentation. The observations that the targets tended to occupy a small fraction of the image and that they tended to be differentially heated led to the hypothesis that the target brightness levels were contained in the "rough" region of the histogram. Roughness, in this context, is the property of the slope of the histogram to change signs frequently. Experiments demonstrated this hypothesis to be correct.

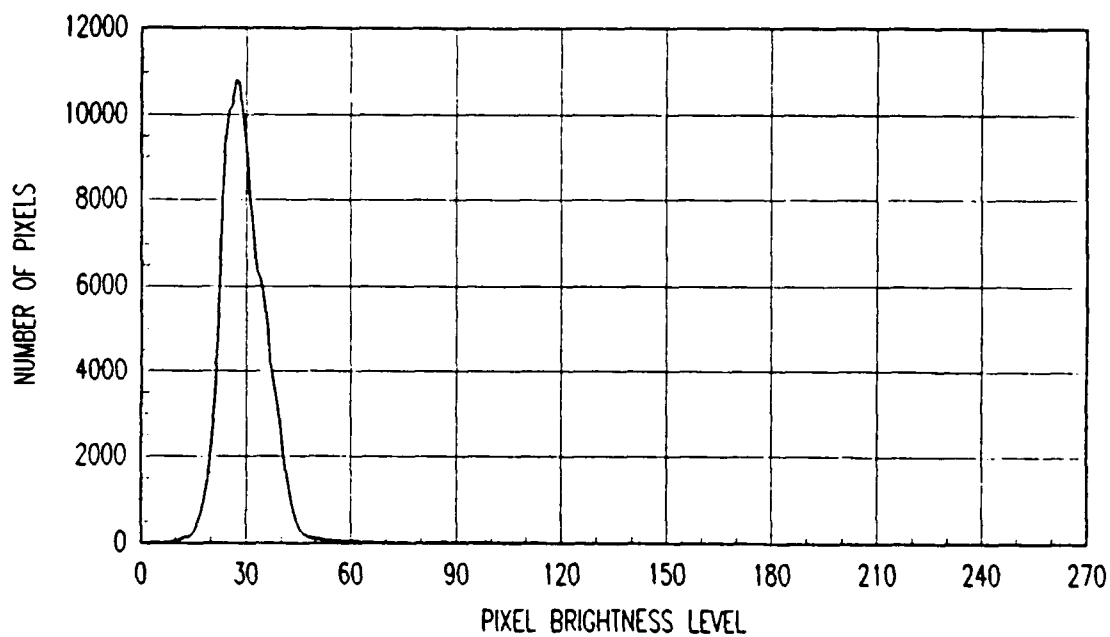


(a)

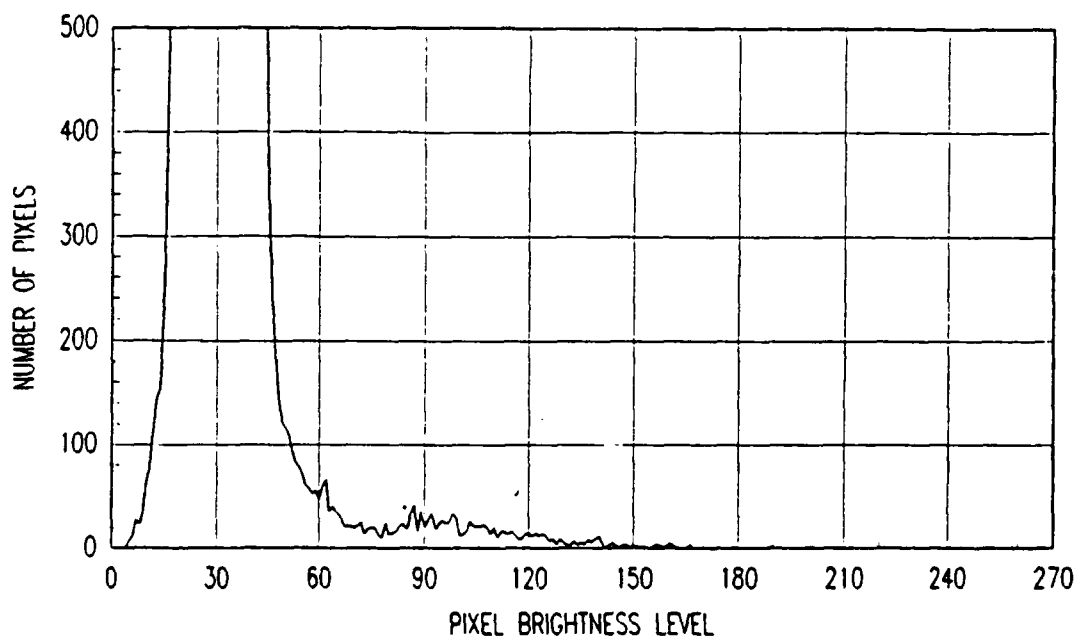


(b)

Figure (3-2). Typical images from the FLIR image data base.



(a)



(b)

Figure (3-3). Histogram of the median filtered version of the images shown in Figure (3-2a): (a) full histogram; (b) vertical axis stretched.

The rule developed to adaptively select the threshold was based on sensing where the rough part of the histogram began. A change in the sign of the slope of the histogram was observed by searching the histogram from the image mean value toward the higher values and noting the first brightness level,  $i$ , at which the following condition was met:

$$H(i+1) > H(i-1): i > \mu_i \quad (3-1)$$

where  $H(i)$  is the histogram value at brightness level  $i$ , and  $\mu_i$  is the mean brightness of the image. Let this level be denoted  $i_1$ . An additional brightness level,  $i_2$ , was obtained by observing the first brightness level at which the following condition was met:

$$(H(i+1) - H(i-1)) > \Delta_H: i > \mu_i \quad (3-2)$$

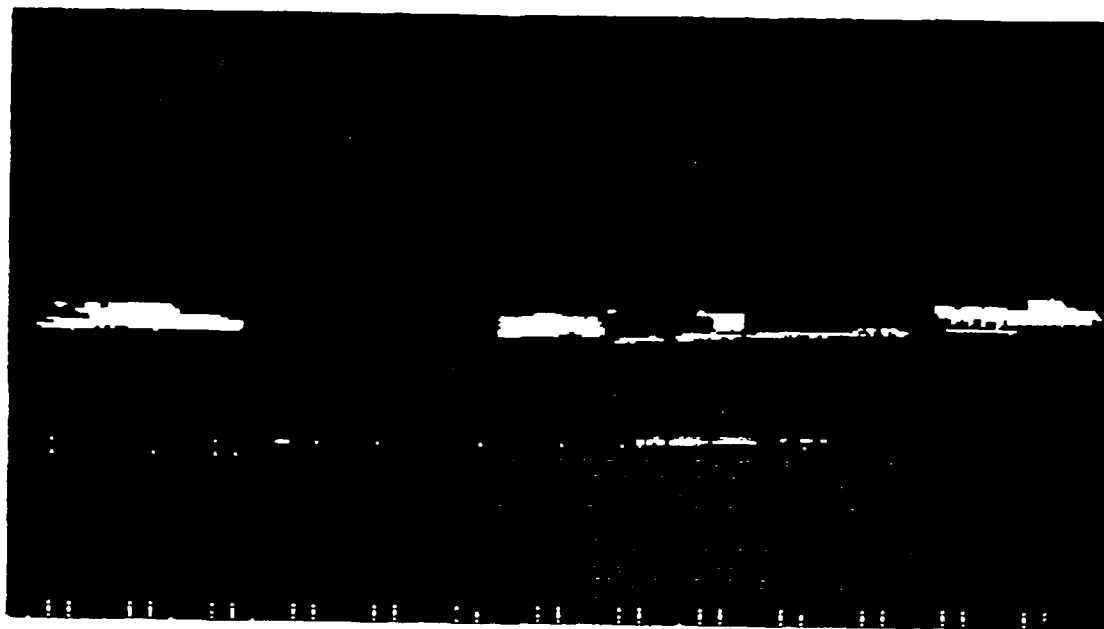
where  $\Delta_H$  is an arbitrary parameter which was set at  $\Delta_H = 15$  for the entire data base. The threshold was then set by the rule:

$$i_{TH} = \frac{(i_1 + i_2)}{2} \quad (3-3)$$

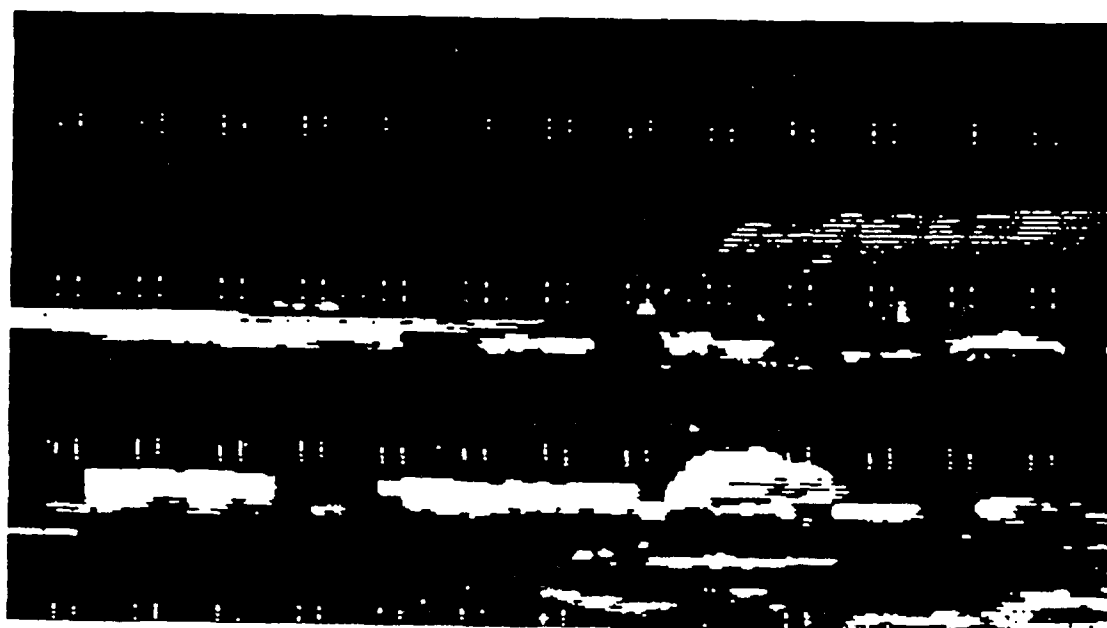
where  $i_{TH}$  is the threshold chosen. This rule selected a threshold of  $i_{TH} = 60$  for the histogram in Figure (3-3). Thresholds between 52 and 75 were selected for images in the data base. A default threshold of  $i_{TH} = 58$  was provided for the rare case when the above rule failed to choose a threshold.

The threshold operation passed a pixel in the median filtered image to the post-threshold image if the brightness of the pixel was greater than the threshold. Pixels in the median filtered image less than or equal to the threshold were set to zero in the post-threshold image.

Binarized versions of the post-threshold images for the images shown in Figure (3-2a) and (3-2b) are shown in Figures (3-4a) and (3-4b), respectively. These results were typical of the data base. Targets were generally retained by the threshold operation while the bulk of the non-target pixels were rejected.



(a)



(b)

Figure (3-4). Post-threshold images for the FLIR images shown in Figure (3-2).



A set of heuristic operations were applied sequentially to the post-threshold image to complete segmentation. In the implementation, the input to each operation was the output of the previous step.

The first heuristic operation was to reject isolated collections of pixels in the post-threshold image which were  $3 \times 3$  pixels, or smaller, in extent. This step eliminated spurious collections of pixels which were far too small to be of interest.

Regions possessing 35 pixels or less were also rejected. This operation resulted in the loss of the targets at 1700 m range in Figure (3-2b), which were viewed with the wide field-of-view of the sensor. However, this loss was acceptable since no suitable range sensor data was available for these targets.

Small dropouts of up to  $3 \times 3$  pixel extent were then filled in all remaining connected regions. Filling such dropouts recovered internal target pixels lost through the threshold and provided well-filled regions for the subsequent processes. Dropouts were filled with the value of the corresponding location in the median filtered image.

Next, a process to eliminate tenuous connections to regions was applied. Tenuous connections to regions were thin strings of pixels, a few pixels wide, attached to larger regions and sometimes connecting two or more regions. This operation was designed to eliminate these connections, rejecting some non-target pixels which passed the threshold.

Each region of sufficiently large vertical extent, defined as seven or more pixels, was contracted by one pixel. The contraction was accomplished by locating every pixel in a region which had a zero-valued pixel as a nearest neighbor, and then setting these pixels to zero. Thread-like connections to regions one or two pixels wide were eliminated by this process. Subregions  $3 \times 3$  pixels or smaller, which were fractured as a result of the contraction operation, were then rejected. The remaining regions were then dilated by one pixel using the reciprocal of the contraction operation.

Tenuous connections to regions with vertical extent of six pixels or less required special attention, since successively contracting and dilating a narrow region typically

resulted in great loss of shape detail. In this case, tenuous connections were fractured by examining the 3x3 pixel nearest neighborhood of each non-zero pixel in the region. If two or fewer non-zero pixels were found in the 3x neighborhood, excluding the center pixel, then the center pixel was set to zero.

Regions possessing a length-to-width ratio greater than a specified upper bound were rejected next. This was reasonable, since the subsequent steps sought to rejoin regions rather than fracture them. The upper bound on length-to-width ratio was liberally set at 15.0. None of the targets in the data base actually possessed a length-to-width ratio as great as 15.0. However, the FLIR sensor was oversampled in the horizontal dimension by a factor of two (see Appendix A), doubling the length-to-width ratio of the targets. Also, thermal coupling of the lower surfaces of targets to the ground often reduced the vertical extent of the segmented targets, increasing the length-to-width ratio.

An operation to reconnect regions which were inadvertently fractured by the previous steps was then applied. Inadvertent fracturing was an occasional problem, for example, at the point where the cab of a 2.5 ton truck joined the box. This region tended to be dimmer than other parts of the truck, and also tended to possess much less vertical extent than the rest of the target. Hence, both the threshold operation and the operations to eliminate tenuous connections could fracture trucks at this point.

A window of the shape shown in Figure (3-5) was used to recover these pixels. The center pixel in this window was passed over every zero-valued pixel in the current version of the segmented image. If at least one pixel on each side of the window was found to be non-zero, then the center pixel was set to the value of the corresponding position in the median filtered image. While this operation served to reconnect regions, it also sometimes added a few additional pixels to regions which were not inadvertently fractured.

In the final step of the heuristic operations, regions which possessed length-to-width ratios larger or smaller than the range allowed for the targets of interest were rejected.

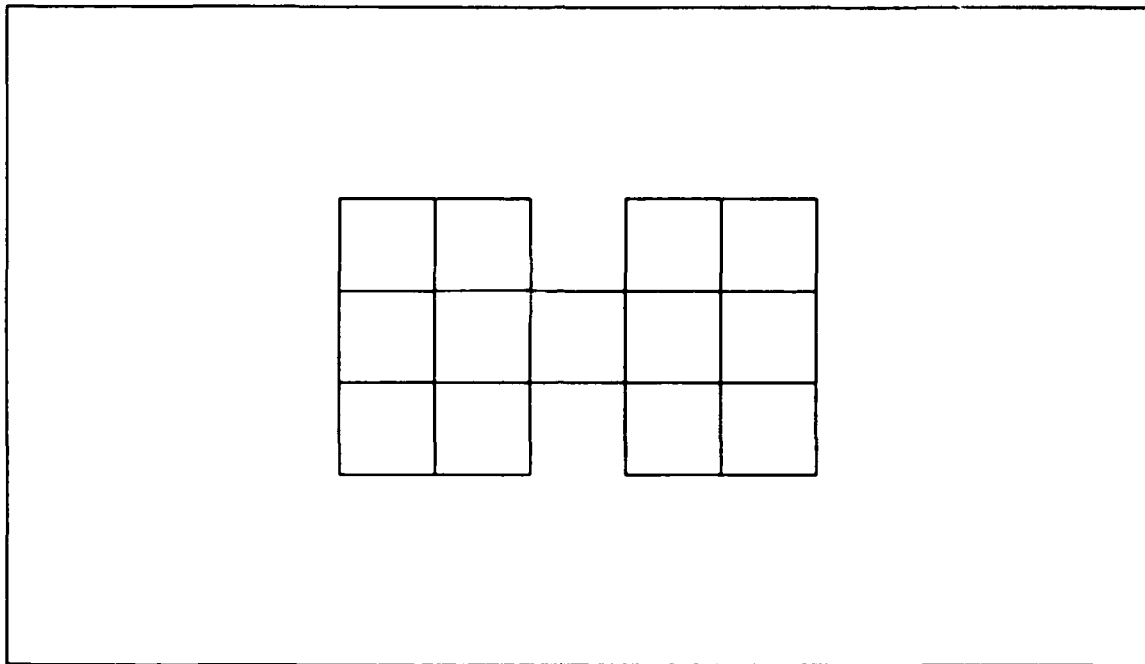


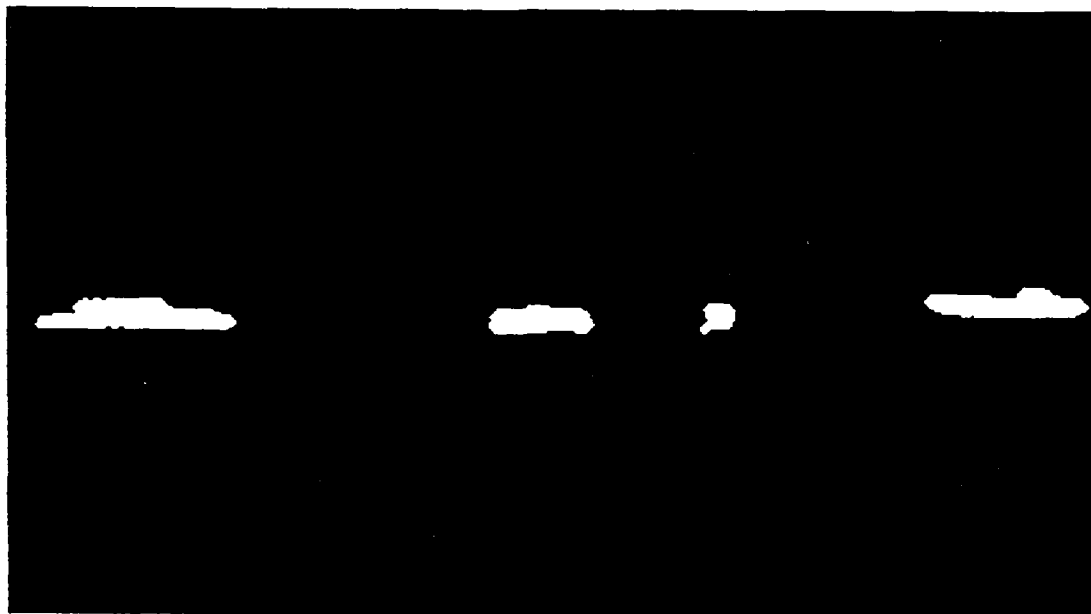
Figure (3-5). Window used in recovery of lost target pixels.

Bounds were set loosely: the lower bound was set at 0.8, and the upper bound was again set at 15.0.

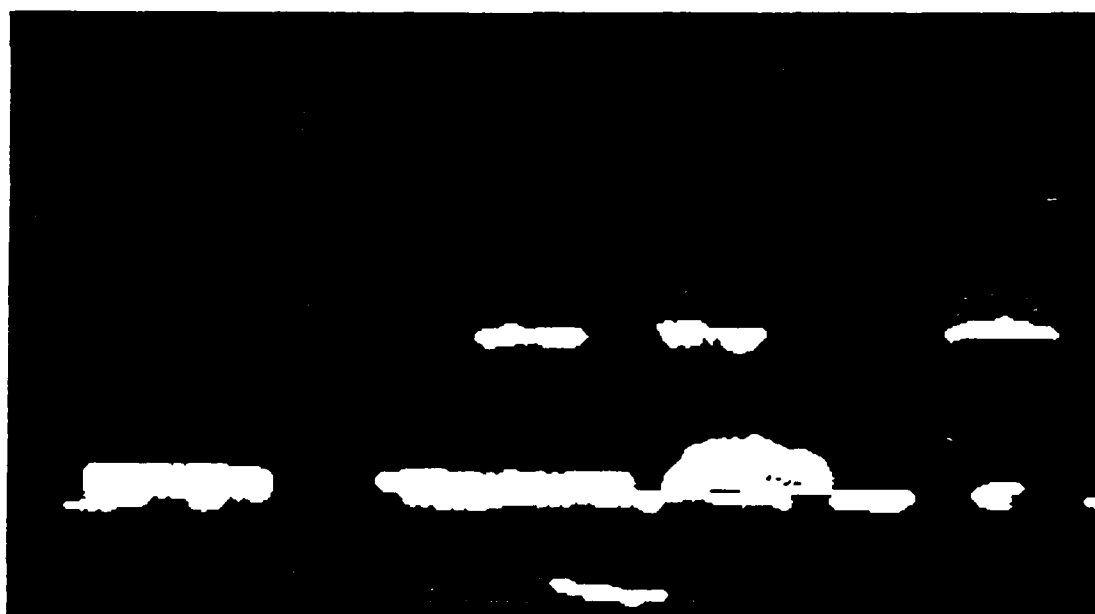
The results of applying these heuristics to the images of Figures (3-4a) and (3-4b) are shown in Figures (3-6a) and (3-6b), respectively. Figures (3-6a) and (3-6b) constitute typical examples of the output of the segmentation algorithm.

Figures (3-6a) and (3-6b) show that the segmentation algorithm was an imperfect selector of targets. In Figure (3-6a) all three targets were passed by segmentation, along with three non-target regions, including the target board. In Figure (3-6b) two trucks and one tank were passed, along with several non-target regions. The tank which was merged with the background in brightness in Figure (3-2b) was lost during segmentation due to its proximity with an equivalently bright section of the background.

Figures (3-6a) and (3-6b) illustrate the need for further processing of segmented images before making target detection declarations. The post-segmentation target detection problem was that of separating segmented target regions from segmented non-target



(a)



(b)

Figure (3-6). Final segmented version of the images shown in Figure (3-2).

regions.

While the algorithm described above was closely tuned to the sensor and data base at hand, the general philosophy of segmentation embodied by this algorithm may be applicable to other problems. The initial operation, in this case a threshold, sought to reject as many background pixels as possible, while both fracturing the targets from the background and keeping as many target pixels as possible. Succeeding steps were designed to eliminate still more unwanted pixels based on insight into the types of image artifacts and targets present. Finally, an effort was made to recover a small number of pixels inadvertently lost during earlier stages of the algorithm.

### **3.3 Algorithm Performance and Scoring Method**

This algorithm was applied to a data base of 97 FLIR images composed of 84 narrow field-of-view images and 13 wide field-of-view images found suitable for multiple sensor research. The data base contained 279 visible targets (not including the targets at 1700 m range imaged with the wide field-of-view), of which 254 were passed by the segmentation algorithm. Thus, targets were passed by segmentation at a rate of 0.910. There were 320 non-target regions passed by the segmentation algorithm. The 254 targets were contained in 230 segmented target regions, for reasons explained below. Thus, the rate of segmented non-target regions per segmented region was  $320/(230 + 320) = 0.582$ . Normalized on a per square degree of image space basis, the false alarm rate was 0.181 per square degree.

Correct target segmentations were scored if a target visible to an observer in the raw FLIR image appeared to be accurately segmented. In several cases in the data base a tank was occluding a jeep, with the result that both the tank and the jeep were segmented as a single connected region. Bounds between the tank and the jeep could not be visibly determined in these cases. For the purposes of computing the segmentation score given above, these instances were scored as two correct target segmentations for two target opportunities.

False segmentations were scored for every region appearing in a segmented image which did not correspond to a target. Normalization of the false segmentation rate on a per segmented region basis provides an estimate of the likelihood that a segmented region did not contain a target. Normalization of this measurement on a per square degree basis yields an estimate of the algorithm performance as a function of the angular size of the image.

### **3.4 Limits of the Algorithm**

Targets must be significantly brighter than the background of the image to be segmented by the present algorithm. The histogram search technique for choosing a threshold mandates this condition for successful segmentation, though the targets may not always meet this requirement. This condition was largely met in the data base. However, in the instances where the target and the background were at nearly the same brightness level, the algorithm typically failed to segment the target.

Also, targets which are closely spaced, or occluding, will typically not be segmented by this algorithm. No operators were developed which would accomplish this task.

### **3.5 Conclusions**

The segmentation algorithm described extracted potential target-bearing regions based on pixel brightness and heuristic operations. This algorithm was found to be quite useful for the purposes of this research project.

The threshold selection technique used is interesting in that though it is not an optimal threshold, it chose an adequate threshold for a large fraction of the images presented. This threshold selection technique may find application in a fielded system having some version of the present algorithm available as a segmentation option when appropriate conditions exist.

Better segmentation heuristics would have contributed to slightly more accurate segmentation of some targets. In particular, an alternative to the method used here for reconnecting inadvertently fractured regions would have been useful if it did not have the effect of blurring some targets. The effect of the blurring induced by this operator was negligible for the multiple sensor target detection work conducted here. However, even a small amount of blurring may affect future target classification work using the segmented images developed here.

## IV. Range Image Segmentation

### 4.0 Introduction

The problem discussed in this chapter is that of automatically extracting regions bearing targets in absolute range images. This process, called segmentation, had as its goal the reliable and accurate extraction of targets from range images, while rejecting as much of the background as possible.

The range sensor used to provide the data base was an active laser radar (Due and Peterson, 1982; Nettleton and Smiley, 1987; Nettleton, 1989). The targets consisted of tactical vehicles at ranges of 860 m to 1700 m. The optical parameters of the sensor and the data collection methodology are described in Appendix A. The data sets and file names from the Army Center for Night Vision and Electro-Optics (CNVEO) June 1987 Multisensor Data Collection (Nettleton and Smiley, 1987) used in the data base are listed in Appendix B. Appendix B also contains a discussion of the criteria used to select FLIR and range image sets for the multiple sensor data base.

The segmentation algorithm developed was based on the observation that the surfaces of tactical targets are reasonably well modeled as collections of small, approximately planar patches of varying orientations. The natural backgrounds surrounding the targets generally did not possess this quality. Hence, a planarity test was found to be well suited for the initial, and most critical step in the segmentation algorithm.

The critical parameter of the segmentation algorithm was a threshold on the absolute error associated with a plane fit to small areas on the Cartesian surface implied by a range image. A good estimate for this parameter was shown to depend upon readily obtainable range imaging and sensing parameters. Surface orientation information was used only to the extent that the computed plane parameters contributed to the absolute



error associated with each plane fit.

The approach taken to segmenting the targets is distinct from other approaches using planar surface extraction for segmentation (Milgram and Bjorklund, 1980; Duda et al, 1979a; Hoffman and Jain, 1987; Besl and Jain, 1988) in that surface orientation was not explicitly used in the segmentation process. This is due to the fact that, while the targets are well-represented as collections of planar patches on a small scale, the orientation of these patches varies widely across real targets. Explicit processing of surface orientation information was found to be unnecessary for this application.

Figures (4-1) and (4-2) provide an illustration of the difficulties associated with using small-scale surface orientation information for segmentation. Figure (4-1a) is a range image of a truck, oriented approximately normally to the sensor beam, at a distance of approximately 1070 m. The standard deviation of range measurements in the raw range image was estimated, using a noise model for the sensor, at approximately 27 cm, and the linear separation of samples across the truck was approximately 5.4 cm. The raw range image was smoothed using a 3×3 pixel median filter followed by a 3×3 pixel averaging filter to produce the image in Figure (4-1a). A modulo 256 computation was applied to every range pixel for display purposes, allowing a range image with dynamic range much greater than eight bits to be shown on an eight bit display. This presentation technique was applied to the range images shown in this dissertation. Displaying images in this manner injects a cyclical appearance into the image which is not present in the raw data. Figure (4-1b) shows a silhouette of the results of applying the segmentation algorithm to this image.

When planes of the parametric form:

$$z = ax + by + \rho_o \quad (4-1)$$

are fit, in the least-squares sense, to the Cartesian coordinates of all 3×3 collections of range image pixels, an estimate of the parameters  $a$ ,  $b$ , and  $\rho_o$  is computed. The parameters  $a$  and  $b$  contain the interesting surface orientation information in this case. Figure

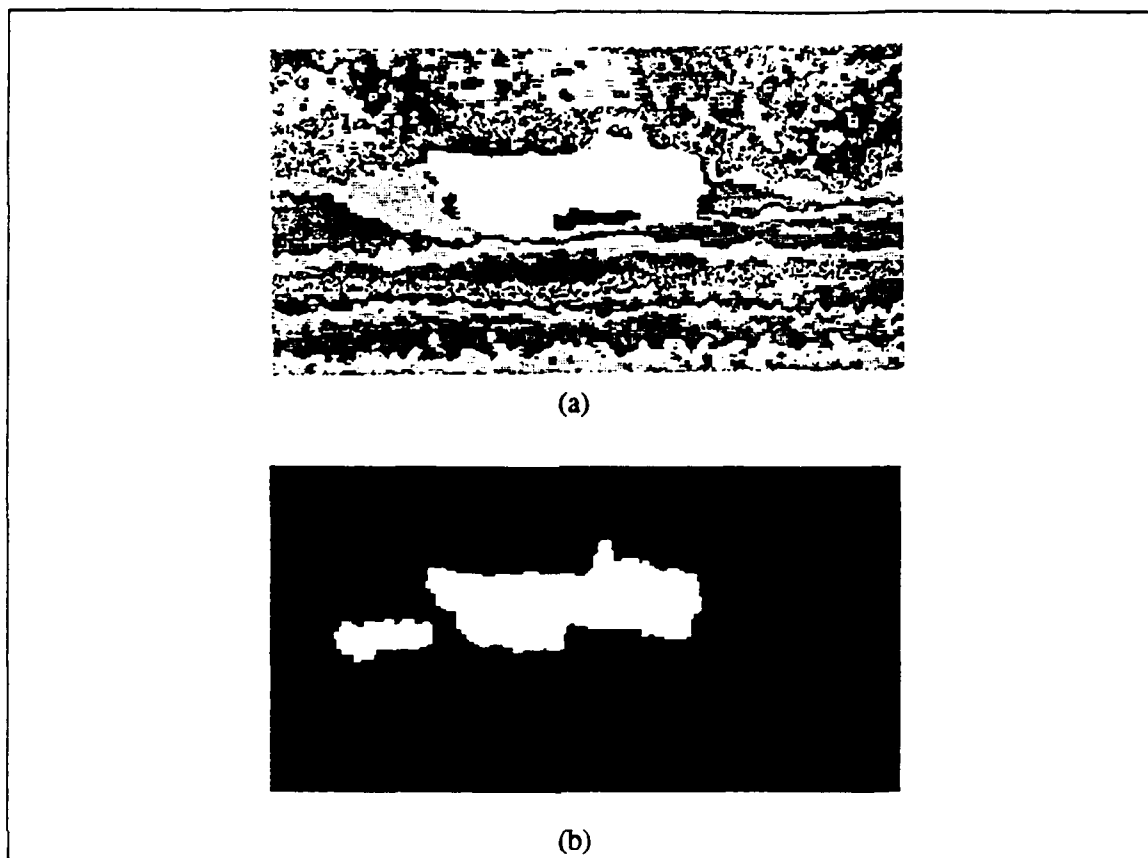
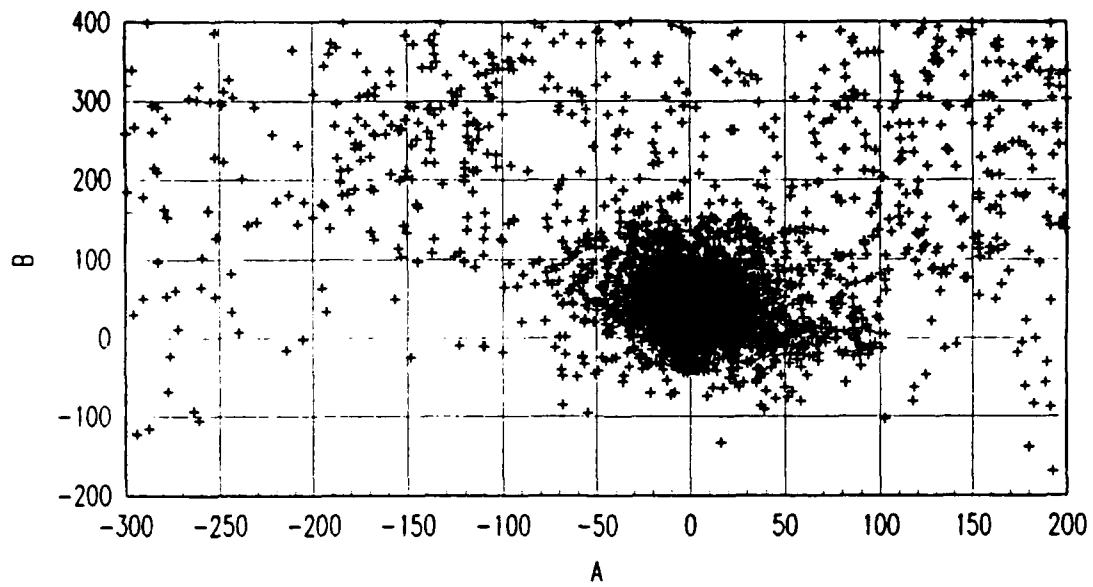
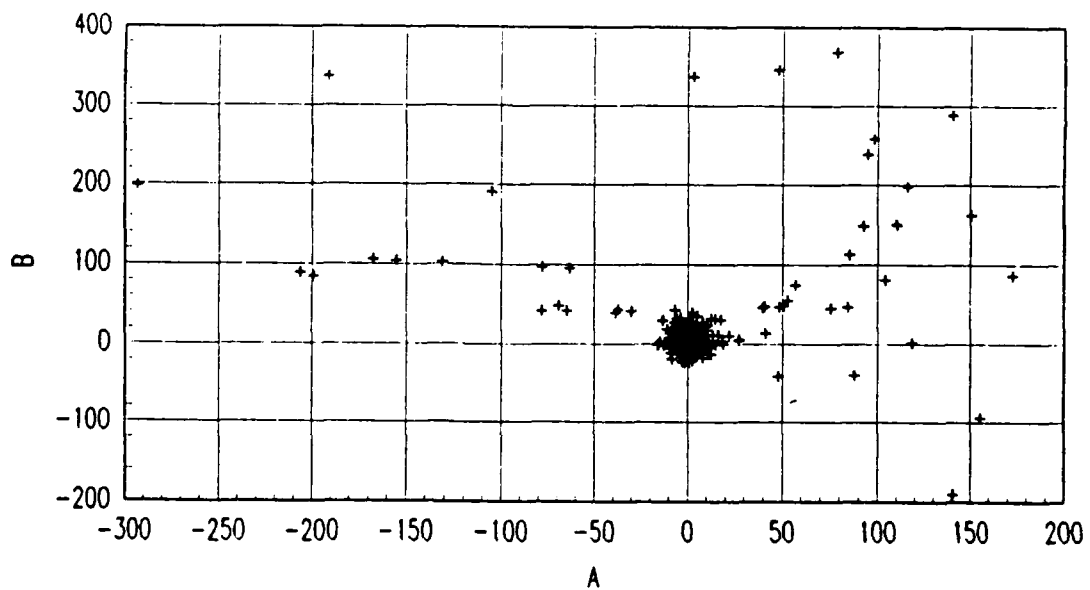


Figure (4-1). (a) Smoothed range image of 2.5 ton truck, broadside view; (b) silhouette of final segmented version of image in Figure (4-1a).

(4-2a) shows a plot of  $b$  versus  $a$  for the pixels in a 128 column by 64 row window of the range image around the truck, with the truck pixels excluded. Figure (4-2b) shows a similar plot, but in this case  $b$  versus  $a$  is plotted for only the truck pixels. Considerable overlap exists between the surface orientations found for the truck and the surface orientations found for the background. Also, the  $a$  and  $b$  parameters for the truck pixels were found to vary over a significant range of values, approximately  $-25 < a < 25$  and  $-40 < b < 40$ . This is a larger spread of values than would be expected if a real truck could be reasonably modeled as a collection of a few truly planar surfaces. Approximately 47% of all pixels in the range image have  $a$  and  $b$  parameters which fall in this region of the parameter space. The planarity test passed approximately 12% of the range pixels, including most of the target pixels, when applied to the image in Figure (4-1a).



(a)



(b)

Figure (4-2). (a) Plane parameters  $b$  vs.  $a$  for 128 column by 64 row window centered on truck, truck pixels excluded; (b)  $b$  vs.  $a$  for truck pixels.

Hence, for the image in Figure (4-1a), the planarity test rejected more of the background than a pixel level segmentation test based on passing pixels in a known section on the  $a-b$  parameter space would have. Segmentation of these images using the information in the  $a-b$  parameter space may be possible, but such an algorithm would need to consider additional information.

The performance of the algorithm developed here shows that explicit processing of surface orientation information was not required to segment the targets in the data base. However, experiments conducted using synthetic range images of planes corrupted with additive, zero-mean Gaussian distributed noise with range-appropriate standard deviation showed that the surface orientations of planes fit to real planar scene regions were computed with good accuracy. The errors in the  $a$  and  $b$  parameters of the planes were found to be approximately zero-mean with standard deviation of less than 1.0 for reasonable operating conditions. Thus, surface orientation information may be useful in identifying objects which have been found using the present technique.

The remainder of this chapter is organized as follows. Pertinent background is presented first. The algorithm and appropriate theoretical considerations regarding the planarity test are then described. This is followed by a discussion of heuristics found useful in the segmentation process. The performance of this algorithm on the database and known limits to applying the algorithm are then presented. Conclusions and comments are made in the final section.

#### **4.1 Background**

Absolute range images provide a measurement of the three-dimensional position of the surface elements in a scene in a coordinate system which has the sensor as its origin. Thus, range image segmentation algorithms typically exploit some property of the surfaces of the objects of interest (Duda et al, 1979a; Milgram and Bjorklund, 1980; Magee et al, 1985; Besl and Jain, 1985; Hofman and Jain, 1987; Besl and Jain, 1988). The philosophy of using some surface property of the targets was adopted in the range image

segmentation algorithm developed.

Recent work in the area of range image segmentation has emphasized processing surface orientation and curvature information (Duda, et al, 1979a; Milgram and Bjorklund, 1980; Magee et al, 1985; Besl and Jain, 1985; Hofman and Jain, 1987; Besl and Jain, 1988). Good success has been reported with these techniques for extracting planar regions in office scenes (Duda et al, 1979a); matching sensed planes to a scene model for position location (Milgram and Bjorklund, 1980); detection of planar, convex, and concave surfaces (Hoffman and Jain, 1987); and extracting higher-order polynomial surfaces (Besl and Jain, 1988). A common theme is the extraction and identification of surfaces in the scene.

The problem of segmenting tactical targets differs from the segmentation problems addressed in the literature. The philosophical difference between the present segmentation approach and previous work is that the 'structure' of the scene, in terms of identifying the types and orientations of the major constituents of the scenes, was a matter of indifference in this project. Rather, a reliable technique was sought for finding targets and accurately partitioning the target pixels from the non-target pixels. It was found that an approach which neglected scene structure and surface orientations in favor of a simple initial test of 'targetness' provided an excellent solution to this problem.

A potential link to the type of surface analysis addressed in the literature exists in the area of analyzing the targets extracted with the present technique. In particular, analysis of the surfaces comprising targets may yield useful insight into their structure, aiding the process of automatically identifying segmented targets. This work was not conducted in this project, but appears promising.

#### **4.2 Segmentation Algorithm**

The segmentation algorithm is shown in block diagram form in Figure (4-3). Raw range images were smoothed using a 3x3 pixel median filter (Gonzalez and Wintz,

1987:p162) followed by a 3x3 pixel averaging filter (Gonzalez and Wintz, 1987:p161). An intermediate image, called the smoothed image, was created by this process. The Cartesian coordinates of each range image pixel were then computed. Planes were fit, in the least-squares sense, to the Cartesian coordinates of all 3x3 pixel regions in the image. The plane parameters computed by the plane fitting routine and the absolute value of the error resulting from the plane fit were associated with the center pixel of the 3x3 region. An intermediate image containing the absolute error values associated with each pixel position, called the error image, was created in this process. A range-dependent error threshold was then applied to the error image such that pixel positions possessing error less than the threshold were passed, while pixel positions possessing error greater than the threshold were rejected, creating an image referred to as the threshold image. Heuristics were applied to the threshold image to reject more non-target pixels and to recover a small number of target pixels. The heuristics included a range-jump test designed to fracture connected regions containing unacceptable jumps in range.

Median filtering reduced the effects of spurious noise which was present in the imagery. The averaging filter further smoothed the image and had the effect of reducing the standard deviation of range measurements by a factor of one third in regions where range changed slowly (Gonzalez and Wintz, 1987:p174).

The Cartesian surface implied by a range image was computed in a coordinate system which had the sensor as its origin, as shown in Figure (4-4). The z-axis of this coordinate system coincided with the boresight of the sensor, and hence the center pixel of the image. The  $(x, y, z)$  position of each range pixel in an image was computed using:

$$x(r, c) = \rho(r, c) \cos \theta_{el}(r, c) \sin \theta_{az}(r, c) \quad (4-2a)$$

$$y(r, c) = \rho(r, c) \sin \theta_{el}(r, c) \quad (4-2b)$$

$$z(r, c) = \rho(r, c) \cos \theta_{el}(r, c) \cos \theta_{az}(r, c) \quad (4-2c)$$

where  $\rho(r, c)$  is the range value at the image (row, column) position  $(r, c)$ , and  $\theta_{az}$  and  $\theta_{el}$  are the angular displacements of the pixel from the center pixel in azimuth and eleva-

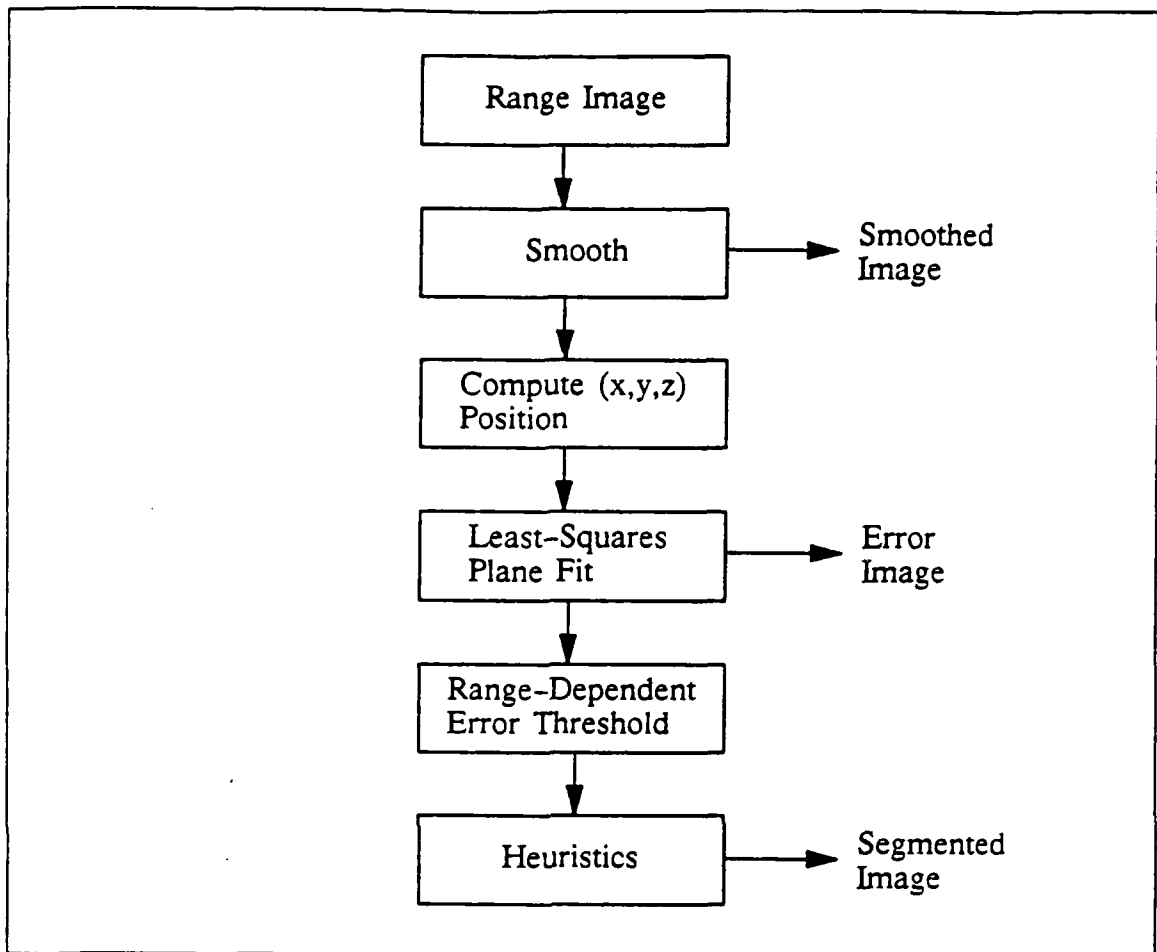


Figure (4-3). Block diagram of range image segmentation algorithm.

tion, respectively.

Planes of the parametric form given by Equation (4-1) were fit, in the least-squares sense, to the  $(x, y, z)$  positions of all  $3 \times 3$  pixel regions in an image, and the absolute error associated with each plane fit was computed. The equations which must be solved to accomplish this fit were derived by applying the standard definition of the least-squares approximation (Burden et al, 1980:p137) to the plane parameterization in Equation (4-1) and setting the partial derivatives with respect to  $a$ ,  $b$ , and  $\rho_o$  equal to zero. The result is that the linear system of equations given by:

$$\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & N_p \end{bmatrix} \begin{bmatrix} a \\ b \\ \rho_o \end{bmatrix} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix} \quad (4-3)$$

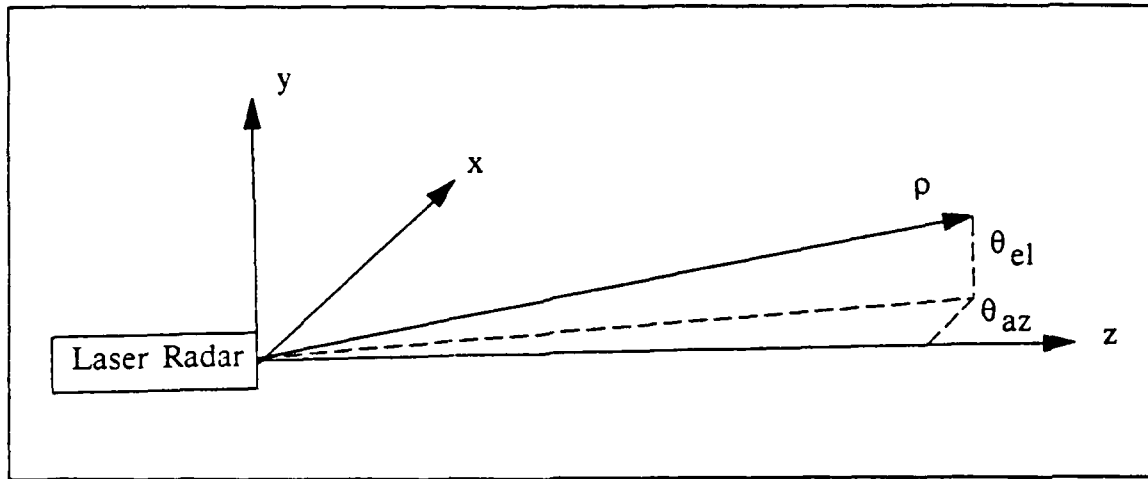


Figure (4-4). Sensing geometry and coordinate system for range image segmentation algorithm.

must be solved for  $a$ ,  $b$ , and  $\rho_o$  each time a plane is fit. All sums in Equation (4-3) are conducted over the nine pixel positions being fit, and  $N_p$  is the number of points being fit, in this case,  $N_p = 9$ . The absolute error associated with a plane fit,  $|e(r,c)|$ , is given by:

$$|e(r,c)| = \sum_{i=-1}^1 \sum_{j=-1}^1 |z(r+i,c+j) - ax(r+i,c+j) - by(r+i,c+j) - \rho_o| \quad (4-4)$$

The values computed for  $|e(r,c)|$  were entered in the  $(r,c)$  position of the error image.

The magnitude of  $z$  was generally several orders of magnitude larger than the magnitude of  $x$  and  $y$  in the data base. Such large disparities in numerical values can cause inaccuracies in the solution of Equation (4-3) (Burden et al, 1980:p12-13). To overcome this difficulty, planes were fit in a translated coordinate system which had the center pixel of the  $3 \times 3$  region as its origin. Thus, planes were actually fit to the Cartesian coordinates given by:

$$x'(r,c) = (x(r,c) - x_o) \quad (4-5a)$$

$$y'(r,c) = (y(r,c) - y_o) \quad (4-5b)$$

$$z'(r,c) = (z(r,c) - z_o) \quad (4-5c)$$



where  $(x_o, y_o, z_o)$  is the  $(x, y, z)$  position of the center pixel in the  $3 \times 3$  region.

The effect of this translation on the parameters of a plane may be explored by assuming a plane has been fit in the translated coordinate system, returning the parameters  $a$ ,  $b$ , and  $\rho_o'$ . Note that the  $a$  and  $b$  parameters are unaffected by translation of the reference coordinate system. However, the  $z$ -intercept is defined in the translated coordinate system. To recover the  $z$ -intercept in the untranslated coordinate system substitute Equation (4-5) into Equation (4-1):

$$\begin{aligned}(z - z_o) &= a(x - x_o) + b(y - y_o) + \rho_o' \\ z &= ax + by + (\rho_o' - ax_o - by_o + z_o)\end{aligned}\quad (4-6)$$

Thus, the  $z$ -intercept in the untranslated coordinate system is given by:

$$\rho_o = \rho_o' - ax_o - by_o + z_o \quad (4-7)$$

The error associated with the plane fit in the translated coordinate system is exactly that which would be obtained from fitting the plane in the untranslated system as seen from:

$$\begin{aligned}|e| &= \sum \sum |(z - z_o) - a(x - x_o) - b(y - y_o) - \rho_o'| \\ &= \sum \sum |z - ax - by - (\rho_o' - ax_o - by_o + z_o)| \\ &= \sum \sum |z - ax - by - \rho_o|\end{aligned}\quad (4-8)$$

where the sums are conducted over the nine pixel positions in the  $3 \times 3$  region. Since the coordinate system translation has no effect on the parameters of interest, the remainder of the mathematical formulation of this algorithm is presented in the untranslated coordinate system.

The planarity test was a threshold operation on the elements of the error image. A range-dependent threshold on  $|e(r, c)|$ ,  $e_T(\rho)$ , discussed in the next section, was computed for every pixel in the range image and applied to create the threshold image,  $T(r, c)$ , using the rule:

$$T(r, c) = \begin{cases} \rho(r, c), & |e(r, c)| \leq e_T(\rho) \\ 0, & \text{otherwise} \end{cases} \quad (4-9)$$

This operation typically rejected a large fraction of the background pixels, while retaining the target pixels. The binarized image shown in Figure (4-5) illustrates the result of applying the error threshold operation to the image in Figure (4-1a). The functional dependence of the error threshold,  $e_T(\rho)$ , on range,  $\rho$ , was of considerable interest. This dependence admits a mathematical analysis, which is presented in the next section.

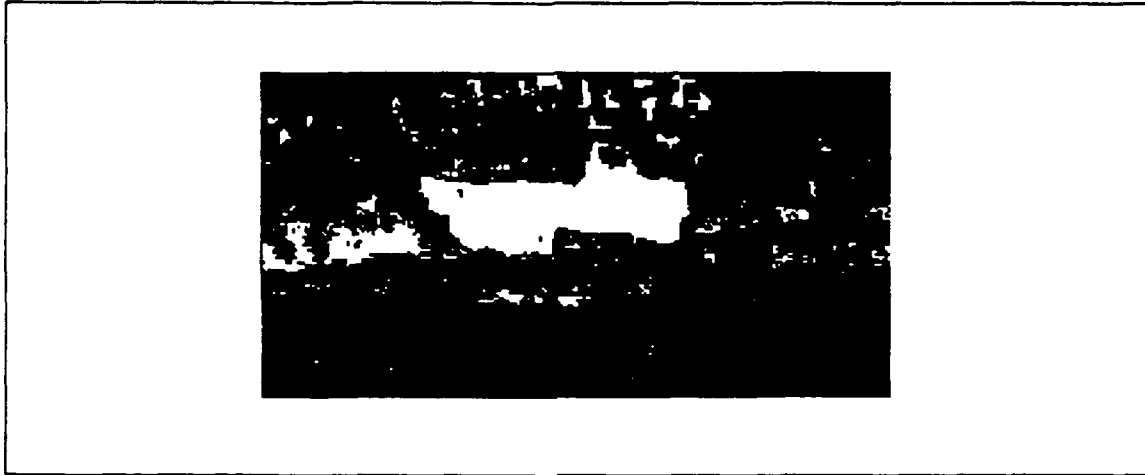


Figure (4-5). Silhouette of image resulting from application of the error threshold to the image in Figure (4-1a).

Equation (4-9) neglects the values of the parameters returned by the plane-fitting algorithm, except as they contribute to computing the absolute error,  $|e(r,c)|$ . Thus, the small-scale planarity of the vehicles was exploited, rather than some property of the surface orientations. It is interesting to note that while man-made vehicles possess small-scale planarity, the natural backgrounds viewed in the data base largely did not possess this quality.

Heuristics were applied to the threshold image,  $T(r,c)$ , as the final step in segmentation. The heuristics were designed to reject more non-target pixels and to recover a small number of target pixels.

### 4.3 Error Threshold Selection

The error threshold discussed above was developed by assuming that the range measurements were corrupted with zero-mean, additive Gaussian noise:

$$\rho_n(r,c) = \rho(r,c) + n(r,c) \quad (4-10)$$

where  $\rho_n(r,c)$  is a noisy range image element,  $\rho(r,c)$  is the actual range of the scene element, and  $n(r,c)$  is the additive noise. The random variable  $n(r,c)$  was assumed to behave as a Gaussian distributed random variable with zero mean and range-dependent variance,  $\sigma_n(\rho)$ .

In the following subsections it is shown that under certain geometrical conditions the mean of  $|e|$ ,  $\mu_e$ , and the standard deviation of  $|e|$ ,  $\sigma_e$ , computed for planar scene regions, are approximated well as functions of only  $\sigma_n(\rho)$ . This is accomplished by exhibiting the geometrical conditions under which the error,  $|e|$ , associated with fitting a plane to the noise-corrupted view of a planar scene region is approximated well as a function of the absolute value of the noise associated with the range measurements for that region,  $|n|$ . A useful rule for choosing the error threshold,  $e_T(\rho)$ , based on this result is presented. Physical considerations for estimating  $\sigma_n(\rho)$  based on sensor parameters and viewing conditions are then discussed.

Numerical experiments were performed to evaluate the accuracy with which plane parameters were computed using noise-corrupted range images of synthetically produced planes. Synthetic range images of planes,  $32 \times 32$  pixels in extent, were corrupted with zero-mean, additive Gaussian noise with range-appropriate standard deviation (see section 4.3.3). The plane parameters in the synthetic image were set at  $a = b = 0$ , and  $\rho_o$  was varied. After corruption, the images were smoothed with a  $3 \times 3$  averaging filter. Planes were fit to the Cartesian coordinates of  $3 \times 3$  pixel regions of these images, and mean and standard deviation of the errors of the computed plane parameters were examined.

Results of the numerical experiments are shown in Table (4-1). In Table (4-1)  $\sigma_e(a)$ ,  $\sigma_e(b)$ , and  $\sigma_e(\rho_o)$  are the standard deviation of the errors on  $a$ ,  $b$ , and  $\rho_o$ , respectively. Errors on the plane parameters  $a$  and  $b$  were found to be approximately zero-mean with standard deviation of less than 0.6 under the conditions of interest. Errors on the parameter  $\rho_o$  were also found to be approximately zero-mean, with standard deviation of less than 1.4 m. Varying the parameters  $a$  and  $b$  up to  $a = b = 40$  did not impact these results. No experiments were conducted to determine where, or if, the errors on plane parameters became large. It was concluded that for the present application, the least-squares plane fitting to noisy range images approximated the actual parameters of the planes observed with sufficient accuracy.

Table (4-1). Error statistics for plane parameters.

| $\rho_o$<br>(m) | $\sigma_e(a)$ | $\sigma_e(b)$ | $\sigma_e(\rho_o)$<br>(m) |
|-----------------|---------------|---------------|---------------------------|
| 800             | 0.43          | 0.51          | 0.50                      |
| 1000            | 0.46          | 0.51          | 0.68                      |
| 1200            | 0.43          | 0.49          | 0.72                      |
| 1400            | 0.47          | 0.53          | 0.95                      |
| 1600            | 0.51          | 0.55          | 1.20                      |
| 1800            | 0.45          | 0.51          | 1.25                      |
| 2000            | 0.47          | 0.56          | 1.34                      |

The standard deviations of the error on  $a$  and  $b$  in Table (4-1) are estimates of the accuracy with which surface normal information may be obtained from noisy range images. The spread of values for  $a$  and  $b$  observed for the broadside truck of Figure (4-1a), shown in Figure (4-2b), may now be interpreted. In particular, the region of  $a$ - $b$  parameter space occupied by the truck pixels, approximately  $-25 \leq a \leq 25$  and  $-40 \leq b \leq 40$ , is seen to be due primarily to variations in the orientations of small surfaces on the truck, rather than as a result of sensor noise.

#### 4.3.1 Derivation of $\mu_e$ and $\sigma_e$ for Planar Scene Regions

Presume that a region in the scene has a spatial extent of  $3 \times 3$  pixels or more, and is planar with parameters given by:

$$z(r, c) = ax(r, c) + by(r, c) + \rho_o \quad (4-11)$$

where  $x(r, c)$ ,  $y(r, c)$ , and  $z(r, c)$  are the actual Cartesian coordinates of the nine range pixels in the  $3 \times 3$  region and  $\rho_o$  is the  $z$ -intercept of the plane. Presume further that the range sensor has observed this region and provided noise-corrupted estimates of the range to this region given by Equation (4-10). The actual Cartesian coordinates of the scene points sampled,  $\rho(r, c)$ , are given by Equations (4-2).

It follows from standard geometric considerations that:

$$\rho(r, c) = [x^2(r, c) + y^2(r, c) + z^2(r, c)]^{1/2} \quad (4-12)$$

and hence:

$$\rho_n(r, c) = [x^2(r, c) + y^2(r, c) + z^2(r, c)]^{1/2} + n(r, c) \quad (4-13)$$

It follows from Equations (4-2) and (4-13) that for the planar region being considered:

$$\rho_n = [(\rho \cos \theta_{el} \sin \theta_{az})^2 + (\rho \sin \theta_{el})^2 + (a \rho \cos \theta_{el} \sin \theta_{az} + b \rho \sin \theta_{el} + \rho_o)^2]^{1/2} + n \quad (4-14)$$

where the  $(r, c)$  dependence of the variables has been suppressed. If either of the following conditions is met:

$$\rho \sin \theta_{az} \approx \rho \sin \theta_{el} \approx 0 \quad (4-15a)$$

$$\rho_o^2 \gg (a \rho \sin \theta_{az})^2, (b \rho \sin \theta_{el})^2 \quad (4-15b)$$

then Equation (4-14) may be simplified to

$$\rho_n(r, c) \approx \rho_o + n(r, c) \quad (4-16)$$

To bound the values for  $a$  and  $b$  under which Equation (4-16) is a good approximation we examine Equation (4-15b). Setting  $\rho_o = \rho$  in Equation (4-15b) yields:

$$\frac{1}{\sin^2 \theta_{az}} \gg a^2, \frac{1}{\sin^2 \theta_{el}} \gg b^2 \quad (4-17)$$

Under the most severe conditions in the data base, specifically, lowest resolution (0.2 mr), at the edges of the largest range image (256 lines by 511 columns), Equation (4-17) yields:

$$|a| \ll 19.5, |b| \ll 39.0 \quad (4-18)$$

As the angular displacement from the center of the image decreases, Equation (4-17) becomes less restrictive. For example, at the highest resolution (0.05 mr) at a point half-way from the center of the largest image to any corner of the image, Equation (4-17) yields:

$$|a| \ll 156.3, |b| \ll 312.5 \quad (4-19)$$

These conditions become progressively less restrictive as the center of the image is approached. It was concluded that the condition in Equation (4-17) was well satisfied in the present case, and did not impose a severe restriction to the approximation in Equation (4-16).

Under the conditions in Equation (4-17),  $|e|$  may be approximated by:

$$\begin{aligned} |e| &= \sum_{i=-1}^1 \sum_{j=-1}^1 |z_n(r+i, c+j) - a_n x_n(r+i, c+j) - b_n y_n(r+i, c+j) - \rho_{o_n}| \\ &\approx \sum_{i=-1}^1 \sum_{j=-1}^1 |\rho_n(r+i, c+j) - \rho_{o_n}| \\ &\approx \sum_{i=-1}^1 \sum_{j=-1}^1 |n(r+i, c+j)| \end{aligned} \quad (4-20)$$

where  $x_n$ ,  $y_n$ , and  $z_n$  are the noise-corrupted Cartesian coordinates obtained from substituting  $\rho_n$  from Equation (4-10) into Equations (4-2), and  $a_n$ ,  $b_n$ , and  $\rho_{o_n}$  are the corrupted plane parameters recovered from fitting planes to the noisy data. Equation (4-20) is a good approximation for pixels with  $a_n$  and  $b_n$  satisfying Equation (4-17), and where  $\rho_{o_n} \approx \rho_o$ . Numerical experiments, discussed previously, showed that  $a_n$  and  $b_n$  were, on average, accurate estimates of  $a$  and  $b$ . Thus, it was concluded that the condition that  $a_n$  and  $b_n$  satisfy Equation (4-17) was not restrictive. The condition that  $\rho_{o_n} \approx \rho_o$  is a less elegant approximation, which was found to be acceptable for this application. The effect

of all approximations made in obtaining Equation (4-20) on the error threshold is discussed in the next section.

The mean of  $|e|$ ,  $\mu_e$ , is computed by making the assumption that the  $n(r,c)$  in Equation (4-20) are statistically independent, hence:

$$\mu_e = 9E\{|n(r,c)|\} \quad (4-21)$$

Based on the previous assumption that  $n(r,c)$  is a zero-mean, Gaussian random variable with standard deviation  $\sigma_n(\rho)$ , the probability density function (PDF) of  $n(r,c)$  is:

$$p_n(\epsilon) = \frac{(2\pi)^{-1/2}}{\sigma_n(\rho)} e^{-\frac{\epsilon^2}{2\sigma_n^2(\rho)}} \quad (4-22)$$

Computation of the PDF of  $|n(r,c)|$ ,  $p_{|n|}(\gamma)$  from  $p_n(\epsilon)$  is a standard problem in the theory of random variables [Papoulis, 1965:p131], and only the result is presented.

$$p_{|n|}(\gamma) = \frac{(2\pi)^{-1/2}}{\sigma_n(\rho)} e^{-\frac{\gamma^2}{2\sigma_n^2(\rho)}} U(\gamma) \quad (4-23)$$

where  $U(\gamma)$  is the unit step function. A numerical value for  $\mu_e$  may be obtained by integrating:

$$\begin{aligned} \mu_e &= 9E\{|n|\} = 9 \int_0^{\infty} \gamma p_{|n|}(\gamma) d\gamma \\ &= 9 \left[ \frac{2}{\pi} \right]^{1/2} \sigma_n(\rho) = 7.181 \sigma_n(\rho) \end{aligned} \quad (4-24)$$

The standard deviation of  $|e|$ ,  $\sigma_e$ , is now derived. This accomplished by exhibiting the variance of  $|e|$ ,  $\sigma_e^2$ , and taking the square root of the result. Define:

$$\sigma_e^2 = E\{|e|^2\} - \mu_e^2 \quad (4-25)$$

Thus, to compute  $\sigma_e^2$  it remains to exhibit  $E\{|e|^2\}$ :

$$E\{|e|^2\} = E\left\{\left(\sum_{i=1}^1 \sum_{j=1}^1 |n(r+i,c+j)|\right)^2\right\} \quad (4-26)$$

To make the succeeding argument more compact, the nine  $n(r, c)$  in Equation (4-20) are now enumerated sequentially from 1 to 9,  $n_1, n_2, \dots, n_9$ . Equation (4-26) may then be expressed:

$$\begin{aligned} E\{|e|^2\} &= E\left\{\left(\sum_{i=1}^9 |n_i|\right)^2\right\} \\ &= E\{|n_1| + |n_2| + \dots + |n_9|\}^2 \\ &= E\left\{\sum_{i=1}^9 |n_i|^2\right\} + E\left\{\sum_{i=1}^9 \sum_{j=1, j \neq i}^9 |n_i| |n_j|\right\} \end{aligned} \quad (4-27)$$

Since the  $n_i$  are assumed to be independent the  $|n_i|^2$  are also independent, hence:

$$E\{|e|^2\} = 9E\{|n_i|^2\} + 72E^2\{|n_i|\} \quad (4-28)$$

The quantity  $E\{|n_i|^2\}$  may be found directly:

$$\begin{aligned} E\{|n_i|^2\} &= \int_{-\infty}^{\infty} \gamma^2 \frac{2(2\pi)^{-1/2}}{\sigma_n(\rho)} e^{-\frac{\gamma^2}{2\sigma_n^2(\rho)}} U(\gamma) d\gamma \\ &= \sigma_n^2(\rho) \end{aligned} \quad (4-29)$$

Thus,  $\sigma_e^2$  is given by:

$$\begin{aligned} \sigma_e^2 &= (9\sigma_n^2(\rho) + 72\left[\frac{2}{\pi}\right]\sigma_n^2(\rho)) - (7.181\sigma_n(\rho))^2 \\ &= 3.270\sigma_n^2(\rho) \end{aligned} \quad (4-30)$$

and the standard deviation of  $|e|$  is given by:

$$\sigma_e = 1.808\sigma_n(\rho) \quad (4-31)$$

#### 4.3.2 Error Threshold

In the previous section estimates for  $\mu_e$  and  $\sigma_e$  were derived for planar scene regions, subject to certain approximations and geometrical constraints on the orientation of the planar surface. For scene regions which were planar and satisfied Equation (4-17) it was expected that the observed values of  $|e|$  would behave well in a statistical sense, possessing an ensemble mean of approximately,  $\mu_e$ , and standard deviation of approxi-



mately,  $\sigma_e$ . Nonplanar scene regions and planar regions violating Equation (4-17) were not expected to provide values for  $|e|$  with the statistics derived in the previous section. In fact,  $|e|$  was often found to be quite large for such regions. Planar regions of the scene which satisfy Equation (4-17) were separated from the nonplanar regions of the scene by using a threshold on  $|e|$ .

Using the approximation in Equation (4-20), it is apparent that  $|e|$  is a random variable formed from the sum of nine random variables. Thus, by the central limit theorem, the PDF of  $|e|$  approaches that of a Gaussian random variable with mean,  $\mu_e$ , given by Equation (4-24), and standard deviation,  $\sigma_e$ , given by Equation (4-31).

The functional form of the error threshold used was:

$$e_T(\rho) = \mu_e + \kappa \sigma_e \quad (4-32)$$

Using the Gaussian approximation of the PDF of  $|e|$ ,  $\kappa = 1.96$  will pass approximately 95% of the pixels resulting from planes in the scene when the threshold of Equation (4-9) is applied [Papoulis, 1965:p 65].

An estimate of the standard deviation of the range measurements,  $\sigma_n(\rho)$ , was required to use Equation (4-32). A model of sensor performance was implemented which predicted this quantity based on estimates of the sensor operating conditions. The estimates of  $\sigma_n(\rho)$  were reduced by a factor of 0.333 to account for the effect of passing a 3x3 averaging filter over the image prior to fitting the planes. Physical considerations for this model are discussed in the next section.

Equation (4-32), with  $\kappa = 1.96$ , was used to set the error threshold for the entire data base. Figure (4-6) shows the range dependence of the error threshold used on the data base. The error threshold used to compute the binarized threshold image in Figure (4-5) was computed in this fashion. The majority of the truck pixels were passed by this operation, as were some smaller collections of non-truck pixels scattered around the image. This result was typical of the algorithm performance on targets at this range.

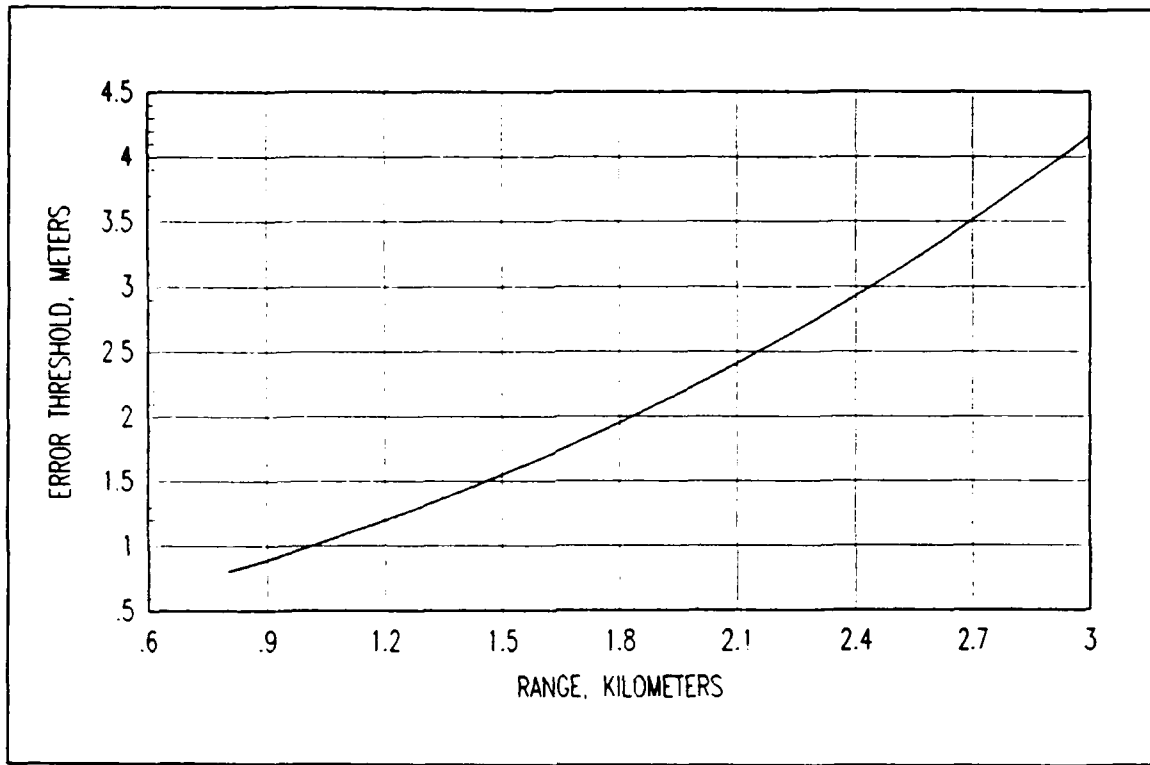


Figure (4-6). Error threshold as a function of range.

Equation (4-32) gives a numerical value for the error threshold of

$$e_T(\rho) = \beta \sigma_n(\rho) \quad (4-33)$$

with  $\beta = 3.57$ , when all contributions to the error threshold are included. Other investigators used empirical means to arrive at a value of  $\beta = 2.5$  for a similar error-of-fit metric (Besl and Jain, 1988:p179).

The cumulative effects of the approximation leading to Equation (4-20) were explored numerically. Synthetic range images of planes,  $32 \times 32$  pixels in extent, with  $a = b = 0$ , were corrupted with zero-mean, additive Gaussian noise with range-appropriate standard deviation (see section 4.3.3). The image was then smoothed with a  $3 \times 3$  averaging filter. Planes were fit to the resulting image, and the following quantities were computed: (1) the mean  $\mu_e$ , and standard deviation,  $\sigma_e$ , of the absolute error associated with the plane fit; and (2) the standard deviation of the range errors in the smoothed

image,  $\sigma_n$ . The error threshold was computed in two ways: (1) using the observed values for  $\mu_e$  and  $\sigma_e$  in Equation (4-32); and (2) using the value obtained for  $\sigma_n$  in Equation (4-33).

| Table (4-2). Comparison of error thresholds. |                      |                    |
|--|----------------------|--------------------|
| $\rho_o$<br>(m)                              | $e_T(actual)$<br>(m) | $e_T(used)$<br>(m) |
| 800  | 0.67                 | 0.79               |
| 1000   | 0.85                 | 0.94               |
| 1200   | 0.99                 | 1.08               |
| 1400   | 1.19                 | 1.37               |
| 1600   | 1.47                 | 1.61               |
| 1800   | 1.66                 | 1.69               |
| 2000   | 1.66                 | 2.07               |

Results of this experiment are shown in Table (4-2). In Table (4-2),  $e_T(actual)$  denotes the error threshold value obtained using the actual values of  $\mu_e$  and  $\sigma_e$  in Equation (4-32);  $e_T(used)$  denotes the error threshold value which would have been used in the segmentation algorithm at the given range, obtained by using  $\sigma_n$  in Equation (4-33). Table (4-2) shows that the method for estimating  $e_T(\rho)$  developed here consistently over-estimated the actual value which would have been obtained had precise values for  $\mu_e$  and  $\sigma_e$  been available. The average magnitude of the over-estimate in Table (4-2) is 12.7%. This was an acceptable result, since the values for  $|e|$  observed for non-planar scene regions were, on average, much larger than the values for  $|e|$  observed for planar scene regions.

#### 4.3.3 Physical Considerations for $\sigma_n(\rho)$

The sensor used to collect the data base was a laser radar which used heterodyne detection. The sensitivity and signal-to-noise ratio performance of such systems, and their impact on  $\sigma_n(\rho)$  is known (Due and Peterson, 1982). A detailed discussion of this topic is beyond the scope of this dissertation. The functional form of  $\sigma_n(\rho)$  and the

important parameters affecting this quantity are now discussed.

For systems of the type used here,  $\sigma_n(\rho)$  depends on the signal-to-noise ratio presented to the signal processor, with the functional form:

$$\sigma_n(\rho) = \frac{v_c}{8\pi f_m} [\delta M (S/N)_P]^{-1/2} \quad (4-34)$$

where  $v_c$  is the speed of light;  $f_m$  is the modulation frequency,  $f_m = 8$  MHz;  $\delta$  is a loss factor  $\delta \approx -8$  dB;  $M$  is the number of samples integrated in the receiver,  $M = 2$ ; and  $(S/N)_P$  is the ratio of signal power to noise power output by the detector (Nettleton, 1989). Numerical values for  $(S/N)_P$  are given by:

$$(S/N)_P = \frac{\eta P (\alpha/\pi)}{h\nu B} \times \frac{\lambda R_c \eta_a}{R_c^2 + \rho^2} \quad (4-35)$$

where  $\eta$  is the system efficiency, with a maximum value of  $\eta = -28.5$  dB;  $P$  is the transmitted power,  $P = 7.5$  W;  $\alpha$  is the reflectivity;  $h\nu$  is the photon energy;  $\lambda$  is the wavelength,  $\lambda = 10.6$  micrometers;  $R_c$  is a constant distance,  $R_c = 561.66$  m;  $\rho$  is the range to the scene element; and  $\eta_a$  is the atmospheric losses (Nettleton, 1989). Combining Equations (4-34) and (4-35), and substituting constants yields:

$$\sigma_n(\rho) = 6.797 \times 10^{-7} \left[ \frac{\eta_a \eta \alpha}{R_c^2 + \rho^2} \right]^{-1/2} \quad (4-36)$$

when  $\sigma_n(\rho)$ ,  $R_c$ , and  $\rho$  are expressed in meters.

Estimates of system efficiency,  $\eta$ , atmospheric transmission,  $\eta_a$ , and nominal scene reflectivity,  $\alpha$  were required to obtain numerical values from Equation (4-36). Values used were:  $\eta = -32$  dB,  $\eta_a = 1.0$  dB/km, and  $\alpha = 0.02$ . These values yield the dependence of  $\sigma_n(\rho)$  upon range shown in Figure (4-7). This estimate of the range dependence of  $\sigma_n(\rho)$  was used as the input to Equation (4-33) to obtain an error threshold for the entire data base. Excellent results were obtained.

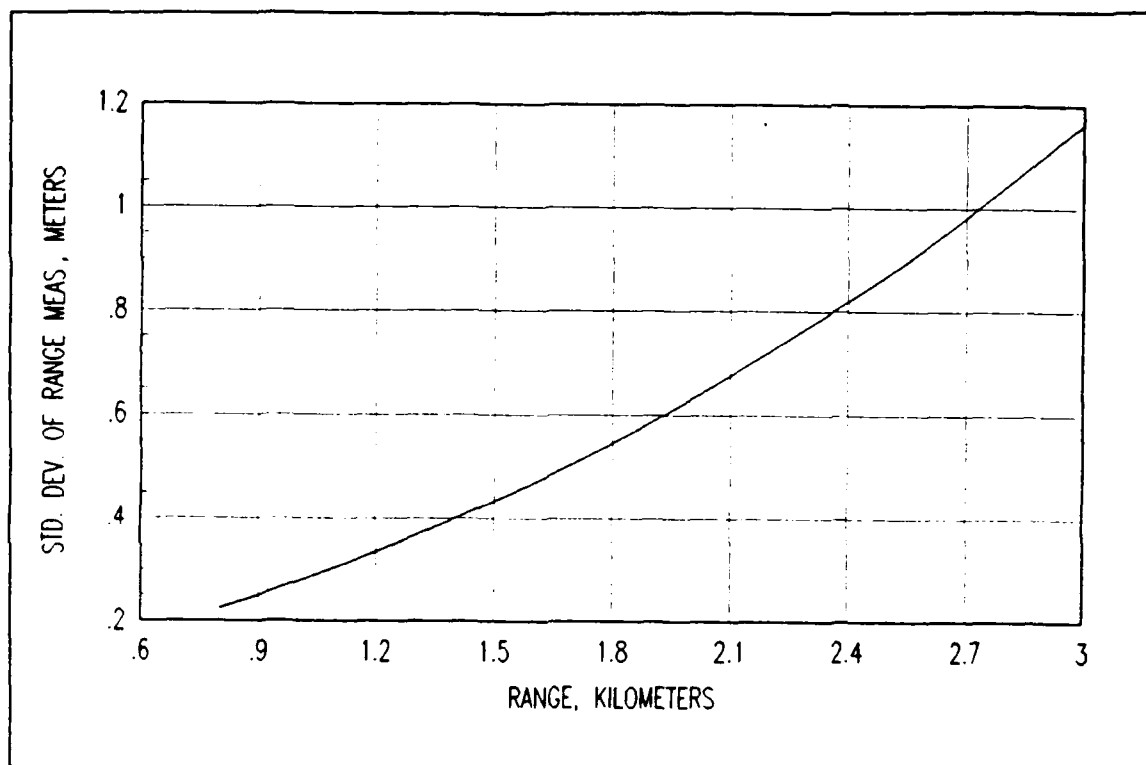


Figure (4-7). Standard deviation of range measurements as a function of range.

#### 4.4 Heuristics for Segmenting Range Images

Figure (4-5) illustrates that additional steps were required to complete segmentation after the error threshold was applied. Heuristics were used to accomplish this task. The heuristics used had three objectives: fracturing connected regions in the threshold image which contained large range jumps, recovering object pixels lost during the error threshold, and rejection of connected regions which were too large or too small to be objects of interest.

Fracturing connected regions in the threshold image,  $T(r,c)$ , which contained unacceptably large range jumps was required to account for the possibility that regions passing the error threshold were connected in the threshold image, but were in fact separated by a large step in range. A range-jump test was applied to all non-zero pixels in  $T(r,c)$ . This test had the form:

$$\begin{aligned}
m &= \max |T(r+i, c+j) - T(r+k, c+l)| \\
&\quad -1 \leq i, j, k, l \leq 1, i \neq k, j \neq l, T(\bullet) \neq 0 \\
T(r, c) &= \begin{cases} 0, & m > T_{RJ} \\ T(r, c), & \text{otherwise} \end{cases} \quad (4-37)
\end{aligned}$$

In words, Equation (4-37) means that for  $3 \times 3$  regions around all non-zero pixels in  $T(r, c)$  the maximum delta range for the non-zero pixels in the region,  $m$ , was computed. If  $m$  was greater than the range-jump threshold,  $T_{RJ}$ , then the center pixel of the  $3 \times 3$  region was set to zero; otherwise, the pixel was not affected. A useful rule for selecting  $T_{RJ}$  was:

$$T_{RJ} = p \theta_{res} \rho_m + \alpha \sigma_n(\rho) \quad (4-38)$$

where  $\theta_{res}$  is the resolution of the sensor,  $\rho_m$  is the mean range of the region, and  $p$  and  $\alpha$  are multiplicative constants. Values of  $p = 5.0$  and  $\alpha = 1.0$  were used for the data base.

Where large range jumps existed in the image, particularly at the boundaries of the objects and the background, planes fit poorly. As a result, the edge pixels of the objects passed by the preceding steps were lost through the error threshold operation. All regions remaining after application of the preceding steps were dilated by one pixel to account for this process.

Regions which were too large or too small were rejected using knowledge of the absolute size of the objects of interest. The angular extent of each region remaining after application of the range-jump test was compared to the maximum and minimum possible angular extent of the family of objects of interest, at the mean range of the region being examined. Regions failing this test were set to zero. This test neglected the orientation of the targets. However, this test was found to be quite useful in rejecting non-target regions.

An additional demonstration of the algorithm is provided in Figures (4-8). Figure (4-8a) is the smoothed range image of a 2.5 ton truck viewed in the front-passenger side

aspect. Figure (4-8b) is the silhouette of the threshold image for Figure (4-8a). Figure (4-8c) is the silhouette of the final segmented version of the image in Figure (4-8a). The range image segmentation system was an imperfect selector of targets, as nontarget regions remained after segmentation was complete. However, the reliability and accuracy with which targets were typically extracted made this algorithm very useful for the present work.

#### **4.5 Algorithm Performance and Scoring**

This algorithm was applied to a data base of 57 range images found suitable for multiple sensor research. The data base contained 137 visible targets, of which 121 were passed by segmentation. Thus, targets were passed by segmentation at a rate of 0.88. The 121 targets were contained in 124 segmented target regions, for reasons explained below. There were 276 non-target regions passed by the segmentation algorithm. Thus, the rate of segmented non-target regions per segmented region was  $276/(124 + 276) = 0.690$ . Normalized on a per square degree of scene space basis, the false segmentation rate was 1.613 per square degree.

Correct target segmentations were scored if a target visible to an observer in the smoothed range image appeared in the segmented image. Occasionally, targets were fractured into two distinct regions due to high noise. Fractured targets were scored as one correct segmentation for one segmentation opportunity.

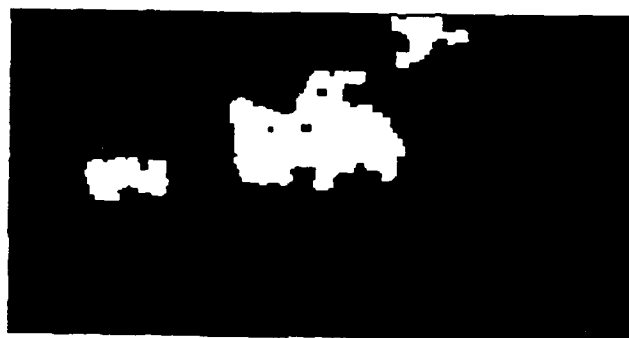
False segmentations were scored for every region appearing in a segmented image which did not correspond to a target. Normalization of the false segmentation rate on a per segmented region basis provides an estimate of the likelihood that a segmented region did not contain a target. Normalization of this measurement on a per square degree basis yields an estimate of the algorithm performance as a function of the angular size of the image.



(a)



(b)



(c)

**Figure (4-8).** (a) Smoothed range image of 2.5 ton truck, front-passenger side view; (b) threshold image for Figure (4-8a); (c) final segmented version of image in Figure (4-8a).



#### 4.6 Limits of the Algorithm

High concentrations of spurious noise on the targets was the primary cause of segmentation failures. Where the spurious noise on a target was too dense for the smoothing scheme to overcome, the error associated with fitting planes became large. Target pixels, and occasionally entire targets, were lost through the error threshold as a result. Better range sensors, with reduced spurious noise, will result in better segmentation performance using this algorithm.

Two additional limits on extending the present algorithm are known. First, the targets to be found must be reasonably approximated as planes on the scale of the area subtended by  $3 \times 3$  pixel regions at the ranges of interest. Second, the target surfaces must not be corrupted by devices which obscure their surfaces. Both of these factors contribute to increasing the error associated with fitting planes to target surfaces, and hence to reducing the number of target pixels passed by the error threshold.

#### 4.7 Conclusions

A range image segmentation algorithm was described which extracted objects composed of small planar regions in the presence of additive, zero-mean Gaussian noise corrupting the range measurements. The segmentation performance obtained from this algorithm was found to satisfy the needs of this project.

Segmentation was accomplished through use of a planarity test which examined the absolute value of the error,  $|e|$ , associated with each plane fit. A range-dependent threshold on  $|e|$ ,  $e_T(\rho)$ , was developed to accomplish this test. It was shown that for planar scene regions under certain reasonable geometrical conditions, described in Equation (4-17), the mean of  $|e|$ ,  $\mu_e$ , and the standard deviation of  $|e|$ ,  $\sigma_e$ , are well-approximated by functions of the standard deviation of the range measurements,  $\sigma_n(\rho)$ . These results, contained in Equations (4-24) and (4-31), were very useful since  $\sigma_n(\rho)$  may be estimated from sensor parameters and operating conditions.

The Gaussian approximation for the PDF of  $|e|$  was used to develop a rule for selecting  $e_T(\rho)$  based on a criterion of selecting approximately 95% of the pixels resulting from viewing planar regions in the scene. This rule is exhibited in Equation (4-32). A model for estimating  $\sigma_n(\rho)$  based on physical considerations was developed and included in the segmentation algorithm. Numerical experiments showed that the approximations leading to the functional expression for the error threshold given in Equation (4-33) were acceptable for the present application.

The segmentation algorithm selectively extracted regions composed of small areas which reasonably approximated planes. The orientation of the planes was neglected by this algorithm. It was concluded that the targets possessed the property of small-scale planarity, while only small portions of the background had this property.

Surface normal information was not explicitly used by the algorithm, and was not required for the segmentation process. However, numerical experiments showed that the surface orientation information was extracted with sufficient accuracy, on average, to allow useful analysis of the surface orientations of segmented objects. Surface orientation information may prove useful for identifying segmented objects.

## **V. Features and Geometric Registration**

### **5.0 Introduction**

The problems addressed in this chapter are the determination and measurement of single and multiple sensor features, and geometric registration of segmented regions between the images. Sensor-dependent features and a novel multiple sensor feature, called the correspondence feature, were used to estimate the class, target or non-target, of segmented regions in FLIR and range images. Geometric registration was required to measure the multiple sensor correspondence feature, since the imagery was not pixel registered.

It has been noted that the choice of features and the design of the classifier are often, in practice, inseparable processes (Fukunaga, 1972:4; Devijver and Kittler, 1982:192-193). This philosophy was adopted in this research. Though it is convenient to discuss the features and the classifier in separate chapters, it is impossible to discuss the selection of features without discussing the classifier. Thus, references to the classification algorithm will appear in this chapter. The classifier is discussed in Chapter VI, and the reader is referred to that chapter for questions regarding the classifier design.

The set of sensor-dependent features initially considered were chosen based on exploiting properties of the targets compared to non-targets as viewed with a given sensing mode. FLIR image features were based on pixel brightness and gross shape. Range image features were based on size, gross shape, and distance. A selection process was applied to the initial set of features to select a subset for use in the classifier. The selection was based on minimizing the probability of error which would arise from using a single feature in the classifier.

A novel multiple sensor feature, called the correspondence feature, was developed

for use in a multiple sensor environment. This feature was designed to exploit the observation that targets appear in the same space in both types of image, while segmented non-targets do not tend to behave in this manner. The correspondence feature was used to add information to the multiple sensor class-estimation processes based on a directed search of areas in one sensor image based on regional cues from the other sensor image. This search, conducted at the pixel level, was used to estimate whether a target was present in the cued region. It was not necessary for targets to be segmented by both sensors for the correspondence feature to be useful in detecting targets: as part of the correspondence feature computation the initial segmentation criterion in the cued region was reevaluated to test the hypothesis that a target may have been present, but was lost due to high noise or other segmentation problems. The correspondence feature proved to provide a very powerful piece of information to the target/non-target discrimination process.

Accurate geometric registration between the images was required to obtain the correspondence feature. Pixel registration was not required to measure the correspondence feature, but lack of pixel registration was a factor in its development. The correspondence feature was a region-based feature in the sense that it measured properties of a cued region, rather than making a measurement requiring pixel-to-pixel registration between the images. Allowances were made for small mis-registrations between the images in the criteria for assigning the various values of the correspondence feature.

The remainder of this chapter is organized as follows. Background to the problem of feature selection is presented in the next section. The criterion used to select single sensor features from the larger set of features initially considered, and the features selected are then discussed. This is followed by discussions of the geometric registration technique and the multiple sensor correspondence feature. Conclusions and comments are made in the final section of this chapter.

## **5.1 Background**

In pattern recognition the term feature is used to define a measurement which is made on input patterns which contains information useful for distinguishing the various patterns. Unfortunately, there is very little theory to guide the choice of features (Devijver and Kittler, 1982:15). In choosing features, consideration must be given to the physics of the sensor, the nature and complexity of the classification problem, and measurements to demonstrate that the features selected separate the classes. For the present problem, feature selection involved choosing a set of features which showed good class separation (Fukunaga, 1972:258; Devijver and Kittler, 1982:15).

Optimal approaches for choosing the 'best' subset of a larger set of features have been demonstrated (Devijver and Kittler, 1982:204-205). An example of an optimal search algorithm is the "branch and bound" algorithm (Devijver and Kittler, 1982:207-214). Such approaches assure the selection of the best set of features, in a minimum error sense, but can involve large amounts of computation for even simple problems (Devijver and Kittler, 1982:204-205).

Suboptimal approaches to choosing the best set of features reduce the computational burden associated with selecting features. The cost associated with this reduction is that a less reliable set of features may be obtained than would be obtained through an optimal search (Devijver and Kittler, 1982:214-216). One example of a suboptimal approach is the "best features" method, in which the individually best features, as evaluated using some performance criterion, are selected (Lewis, 1962:172-173; Devijver and Kittler, 1982:215-216).

## **5.2 Feature Selection Method**

Several features were initially considered for each type of sensor image. This initial set of features was chosen based on an understanding of the sensor physics, and an assessment of the characteristics of segmented targets when compared to segmented

non-targets. Features which were insensitive to small changes in the pixels present in segmented target regions were used; for example, the length-to-width ratio of segmented regions. Local background, as used below, refers to a rectangular window 50% larger in both length and width than a rectangular box just holding the segmented region, which is centered on the segmented region, and which excludes the pixels in the segmented region.

Nine features for FLIR images were considered: (1) the standard deviation of the brightness levels in a segmented region, called the pixel standard deviation; (2) the difference between the mean brightness level of a segmented region and the mean brightness of the local background, called the difference of the means; (3) the maximum pixel value present in a segmented region, called the maximum pixel value; (4) the ratio of the number of pixels in a region to the number of pixels in a rectangular box just holding the region, called the compactness; (5) the ratio of the number of edge pixels to the number of pixels in a region, called the complexity (Rosenfeld and Kak, 1982:265); (6) the ratio of the difference between the mean brightness of a region and the mean brightness of the local background to the sum of these two mean brightnesses, called the contrast of the means; (7) the internal contrast of a segmented region, called the contrast; (8) the ratio of the number of pixels within 10% of the brightness of the brightest pixel in a region to the total number of pixels in the region, called the bright pixel ratio; and (9) the ratio of the horizontal extent of a rectangular box just holding a segmented region to the vertical extent of this box, called the length-to-width ratio.

Eight range image features were considered: (1) the length of a segmented region, computed from the mean range and the horizontal angular subtense of the region, without compensation for the orientation of the region; (2) the height of the region, computed from the vertical angular subtense of the region, without compensating for its orientation; (3) the length-to-width ratio, as computed for FLIR images; (4) compactness, as computed for FLIR images; (5) complexity, as computed for FLIR images; (6) the standard

deviation of the range values in a segmented region, called the pixel standard deviation; (7) the absolute value of the difference between the mean range of a segmented region and the mean range of its immediate background, called the absolute difference of the means; and (8) the absolute value of the difference between the standard deviation of the range measurements for a segmented region and the standard deviation of the range measurements of the pixels in the local background, called the absolute difference of the standard deviations.

All of the features initially considered were computed for every segmented region and stored. Computation of several of the features mentioned above required access to both the segmented image and an earlier version of the images, specifically, the raw FLIR image and the smoothed range image. The processing architecture shown in Figure (2-1) provided for this by saving the raw FLIR image and the smoothed range image in the image memory.

Discrete class-conditioned probability density functions (PDF) were computed for all of the features using a histogram approach with equally spaced bins (Fukunaga, 1972:184-186; Devijver and Kittler, 1982:424-425). The number of bins used in the histograms were obtained empirically: fifteen bins were used for FLIR image features and seven bins were used for range image features. Numerical values for the class-conditioned PDFs were obtained using the relative frequency of occurrence approach (Papoulis, 1965:34):

$$p(f_i(j)|\theta_k) = \frac{n(f_i(j) \text{ and } \theta_k)}{n(\theta_k)} \quad (5-1)$$

where  $p(f_i(j)|\theta_k)$  is a discrete conditional PDF value,  $f_i(j)$  represents the  $i^{th}$  feature having a value in the  $j^{th}$  bin,  $\theta_k$  is the  $k^{th}$  class, and  $n(\cdot)$  represents the number of occurrences observed in the data base. For the present case, the set  $\{\theta_k\}$  was a two member set with  $\theta_1 = \text{target}$ , and  $\theta_2 = \text{non-target}$ . An example of the class-conditioned PDFs computed in this manner is shown in Figure (5-1). Both of the class-conditioned

PDFs of interest,  $p(f_i(j)|target)$  and  $p(f_i(j)|non-target)$ , are shown in Figure (5-1). These are discrete PDFs, though the points are connected in Figure (5-1) to allow the trends to be more easily observed.

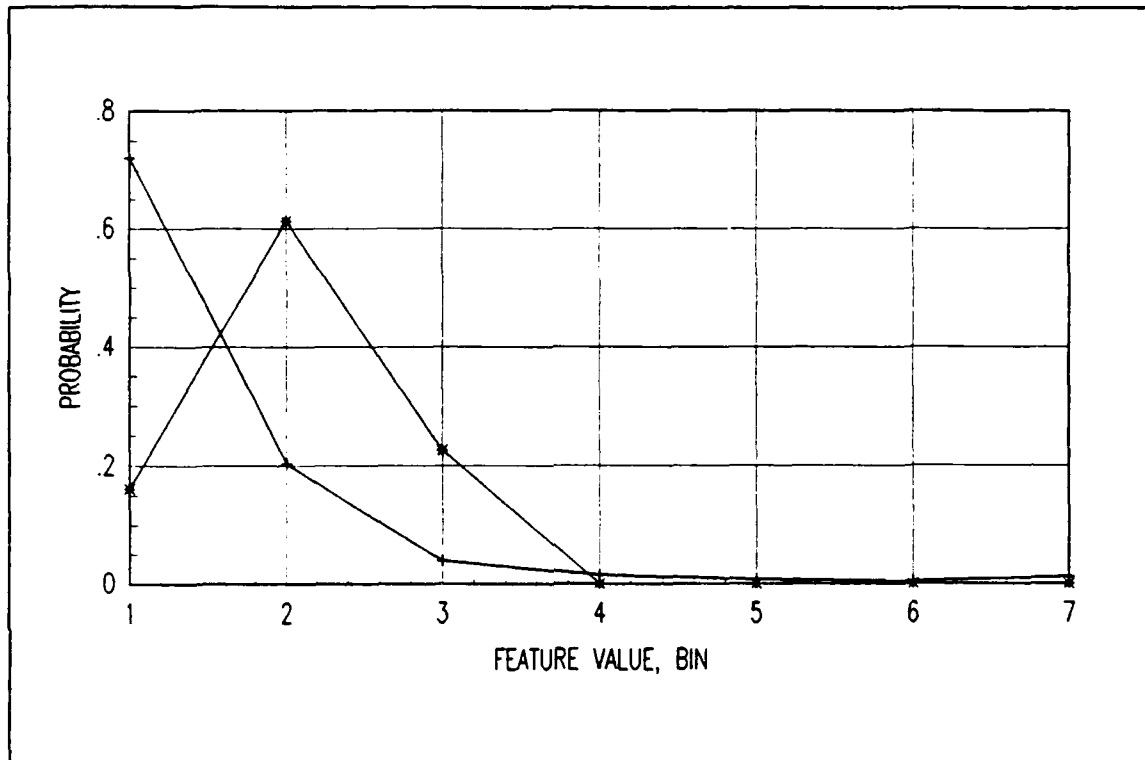


Figure (5-1). Class-conditioned PDFs for the length-to-width ratio feature for range images: \* indicates class = target; + indicates class = non-target.

A subset of the features considered was selected for use in the classifier. This selection was accomplished using the best features approach (Devijver and Kittler, 1982:215-216). The features were rank-ordered using the single feature probability of class estimation error as the ranking criterion.

The probability of class estimation error,  $P_e$ , was computed for each feature using (Melsa and Cohn, 1978:38):

$$\begin{aligned}
 P_e &= P(d=target, t=non-target) + P(d=non-target, t=target) \\
 &= P(d=target|t=non-target)P(non-target) + P(d=non-target|t=target)P(target) \\
 &= 0.5(P(d=target|t=non-target) + P(d=non-target|t=target)) \quad (5-2)
 \end{aligned}$$



where  $P(\cdot)$  is a probability,  $d$  is the single feature class estimate, or decision,  $t$  is the true class of the region, and the factor 0.5 arises from the assumption, made in the classifier, that the classes are, a priori, equally likely. Equation (5-2) was used to compute numerical values for  $P_e$  by noting that the classifier used the Bayesian minimum error decision criterion (Melsa and Cohn, 1978:42), which required that the most likely class be chosen based on each feature observation. Thus, Equation (5-2) may be expressed as:

$$P_e(i) = 0.5 \sum_{j=1}^J p_{\min(k)}(f_i(j) | \theta_k) \quad (5-3)$$

where  $P_e(i)$  is the probability of error for the  $i^{th}$  feature,  $\min(k)$  indicates that the minimum is taken over the two classes, and the summation is taken over the  $J$  bins.

| Table (5-1). Rank-ordered $P_e$ for FLIR image features. |       |
|--|-------|
| Feature  | $P_e$ |
| Complexity   | 0.210 |
| Length-to-width ratio                                    | 0.244 |
| Contrast of the means                                    | 0.259 |
| Maximum pixel value                                      | 0.296 |
| Contrast   | 0.296 |
| Difference of the means                                  | 0.311 |
| Pixel std. dev.  | 0.317 |
| Bright pixel ratio                                       | 0.350 |
| Compactness  | 0.357 |

Single sensor  $P_e(i)$  were computed for the data base. The values obtained were rank-ordered, and are displayed in Tables (5-1) and (5-2) for FLIR and range images, respectively.

The three best features for each type of sensor, as judged by the single sensor probability of class estimation error, were selected to be examined in more detail. Of particular interest was the impact of various combinations of these features on the performance of the classifier. Thus, the FLIR features selected were complexity, length-to-width ratio,

| Table (5-2). Rank-ordered $P_e$ for range image features. |       |
|---|-------|
| Feature   | $P_e$ |
| Length-to-width ratio                                     | 0.202 |
| Abs. difference of the std. dev.                          | 0.262 |
| Complexity  | 0.275 |
| Pixel standard deviation                                  | 0.289 |
| Abs. difference of the means                              | 0.341 |
| Length  | 0.343 |
| Height  | 0.377 |
| Compactness   | 0.398 |

and the contrast of the means. The range features selected were length-to-width ratio, the absolute difference of the standard deviations, and the complexity. The impact of using various combinations of these features in the classifier is discussed in Chapter VI.

### 5.3 Geometric Registration

The need for geometric registration between the sensor images in a multiple sensor system was apparent (Mitiche and Aggarwal, 1986). The data base was not pixel registered. Accurate measurements of the relative positions of the sensors and their pointing angles were also not available. In addition, the resolutions of the two sensors were different (see Appendix A). Thus, pixel-to-pixel registration between the sensor images would have been quite difficult, and was not addressed.

A means of geometrically registering regions between the sensor images was developed. A single pixel, called the common pixel, was chosen in each of a matched pair of images which was taken as originating from the same point in space. Regions were then registered by computing angular displacements from the common pixel for the pixels of interest in one image, and locating the corresponding angular displacements from the common pixel in the other image. This process was called pixel translation. Disparities in the resolutions of the sensors were accommodated in the pixel translation

process. Though pixel translation was a multiple sensor pixel level process, the measurement performed through pixel translation, specifically, computation of the correspondence feature, contained allowances for small errors in locating the common pixel.

The common pixels were obtained from a one-time manual review of the segmented versions of corresponding FLIR and range images. A target which was segmented well in each type of image was selected. The center pixel of a rectangular box just holding the common target in each type of image was selected as the common pixel for each image. The common pixel locations were stored and accessed as required. This method provided an effective means of obtaining geometrical registration from non-pixel registered views of the same scene. Use of this manual technique was merely convenient for the data base, and does not affect the utility of the approach defined for processing non-pixel registered imagery for operational systems. In an operational system the important requirement would be that some means of accurate geometric registration is present. In a well designed operational system the geometric transformation between the images would, most likely, be computed by the sensor positioning systems.

The process of pixel translation was used to locate corresponding positions of segmented regions in one type of image in the other type of image. The notion of one sensor "cueing" a region in the other sensor image arises from this technique. The sensor image which provided the cues for regional searches was called the dominant sensor. The sensor image which was searched was called the non-dominant sensor image.

#### **5.4 Correspondence Feature**

The correspondence feature is a unique multiple sensor feature developed under this project. It was developed to exploit the observation that targets lie in the same space, regardless of which sensor viewed the scene, while segmented non-target regions do not tend to behave in this manner. It was not required that targets be segmented in both types of image for the correspondence feature to provide useful information. A technique for relaxing the segmentation criteria in cued regions was developed to test the possibility

that a target was actually present in a cued region, but was lost during segmentation.

The concepts of the dominant sensor image and the non-dominant sensor image, and the process of pixel translation were important to the computation of the correspondence feature. Segmented regions in the dominant sensor image were used to provide regional cues for searches in the non-dominant sensor image. Pixels in segmented regions of the dominant sensor image were located in the non-dominant sensor image through the process of pixel translation. The correspondence feature value for a segmented region in the dominant sensor image was a function of the properties of the cued pixels in the non-dominant sensor image. Both sensor images were used sequentially as the dominant sensor image so that correspondence feature values were measured for all segmented regions in both types of sensor image.

The correspondence feature had four mutually exclusive possible values: (1) strong correspondence (SC); (2) weak correspondence (WC); (3) weak-weak correspondence (WWC); and (4) no correspondence (NC). The value SC indicated that segmented regions in both types of image occupied very nearly the same space. The value WC indicated that segmented regions in both types of image occupied some of the same space, but not sufficiently well to be declared as a SC. The value WWC indicated that no segmented region in the non-dominant sensor image occupied the cued region sufficiently well to be declared either a SC or a WC, but when the initial segmentation criterion was relaxed in the cued region a sufficient fraction of the cued pixels were found to pass the relaxed initial segmentation test. The value NC was used to indicate that none of the above conditions were met.

A region labeling scheme was used as part of the correspondence feature computation. The labeling scheme accepted a segmented image as input and created an intermediate image called the labeled image,  $L$ , where the pixel locations of connected regions were assigned a new integer value between 1 and  $N_R$ , where  $N_R$  was the number of regions in the segmented image. The approach taken to the region labeling algorithm

was pixel aggregation (Gonzalez and Wintz, 1987:369-373), where the aggregation criterion was occupancy of pixels in the  $3 \times 3$  pixel neighborhood of pixels already identified as members of the  $n^{th}$  region. All pixels in a connected region were assigned the same integer value. This scheme labeled regions in segmented images consistently, and thus provided a tool for identifying segmented regions by a single integer.

The correspondence feature value for a segmented region in the dominant sensor image was computed by applying pixel translation to the pixels in the region and observing some properties of the pixels cued in the non-dominant sensor image. Two properties of the translated pixels were observed: (1) the number of dominant sensor image pixels translated to pixels in the  $m^{th}$  region of the non-dominant sensor image; and (2) the number of dominant sensor image pixels translated to pixels which passed a relaxed version of the critical segmentation criterion for the non-dominant sensor image. For the case of the range image being the non-dominant sensor image, the error threshold (see Chapter IV) was increased by a factor of 1.5. The need to re-examine the error image based on multiple sensor information was the reason for storing the error image in the image memory (see Chapter II). For the case of the FLIR image being the non-dominant sensor image, the brightness threshold (see Chapter III) was reduced by a factor of 0.9.

As the value of the correspondence feature was determined for each segmented region in the dominant sensor image, entries were made in correspondence tables, which are described below. Since the correspondence feature values SC and WC address the joint spatial occupancy of segmented regions in both types of images, the correspondence tables were used to resolve joint spatial occupancy issues in the multiple sensor single decision algorithm. This topic is discussed in Chapter VI.

The value SC had two criteria: (1) at least 55% of the pixels in the dominant sensor image were translated to the same segmented region in the non-dominant sensor image, and if (1) was satisfied: (2) the horizontal angular subtense of the region in the dominant sensor image and the region in the non-dominant sensor image which satisfied (1) were

equal to within  $\pm 20\%$ . The value SC occurred most frequently when targets segmented well in both types of image. When a SC was observed an entry was made in a table, called the strong correspondence table, of the form: the  $m^{th}$  region in the dominant sensor image has a SC with the  $n^{th}$  region in the non-dominant sensor image, which was denoted notationally as  $SC(m) = n$ .

If no SC was found for a region, then the possibility WC was explored. A value of WC was declared for a region in the dominant sensor image if criterion (1) for the value SC was satisfied, but not criterion (2). The value WC occurred most frequently when: (1) targets were partially segmented in the dominant sensor image; or (2) when targets were connected in the non-dominant sensor image, such as the case discussed in Chapter III where a jeep, occluded by a tank, and the tank were segmented as a single region. When a WC was observed an entry was made in a table, called the weak correspondence table, of the form: the  $m^{th}$  region in the dominant sensor image has a WC with the  $n^{th}$  region in the non-dominant sensor image, which was denoted notationally as  $WC(m) = n$ .

If neither a SC nor a WC was found for a region in the dominant sensor image, then the WWC possibility was explored. The WWC accounted for the case where a target was viewed by both sensors, but was lost in one of the images during segmentation. For the case of the FLIR image being the dominant sensor image, a WWC was declared if 30% of the pixels in the FLIR region were translated to range image pixels with error image values of less than 1.5 times the error threshold (see Chapter IV) at the appropriate range. For the case of the range image being the dominant sensor image, a WWC was declared if 70% of the pixels in the range image region were translated to FLIR image pixels with brightness values greater than 0.9 times the brightness threshold (see Chapter III) for the FLIR image. When a WWC was found an entry was made in a table, called the weak correspondence table, of the form: the  $m^{th}$  region in the dominant sensor image has a WWC.

When none of the above correspondence values were observed, the value NC was

declared for the region in the dominant sensor image. The value NC occurred most frequently for segmented non-target regions.

Allowances for small errors in the choice of the common pixel were made in setting the tolerances for the correspondence feature values SC, WC, and WWC. Specifically, the percentages of pixels required to assign the various values for the correspondence feature were developed to allow for the possibility of such errors. Better geometric registration across the data base would probably allow these percentages to be increased. Good performance was obtained with the values described above. Performance is quantified in Chapter VI.

Correspondence feature values were computed for all segmented regions viewed completely by both sensors, and stored. The range images were uniformly completely contained within the FLIR images. Thus, all segmented regions in the range images were also viewed by the corresponding FLIR image, however, the converse was not true. Specifically, many FLIR images contained targets not viewed by the corresponding range image, and segmented FLIR images often contained non-target regions only partially viewed by the range image.

Discrete class-conditioned PDFs were computed for the correspondence feature for the FLIR and range image data bases using Equation (5-1). The PDFs obtained are displayed in Figures (5-2) and (5-3) for FLIR and range images, respectively. Single feature probabilities of error were computed for the correspondence feature using Equation (5-3) as FLIR  $P_e(\text{correspondence feature}) = 0.245$ , and range  $P_e(\text{correspondence feature}) = 0.085$ .

In Figure (5-2) the value for the probability of observing a SC given non-target is shown as 0.0001, the value used in the implementation. In fact, no instances of a non-target region possessing a SC were observed in the data base, giving an observed probability of zero to this possibility. In Bayesian inference processes a probability of zero

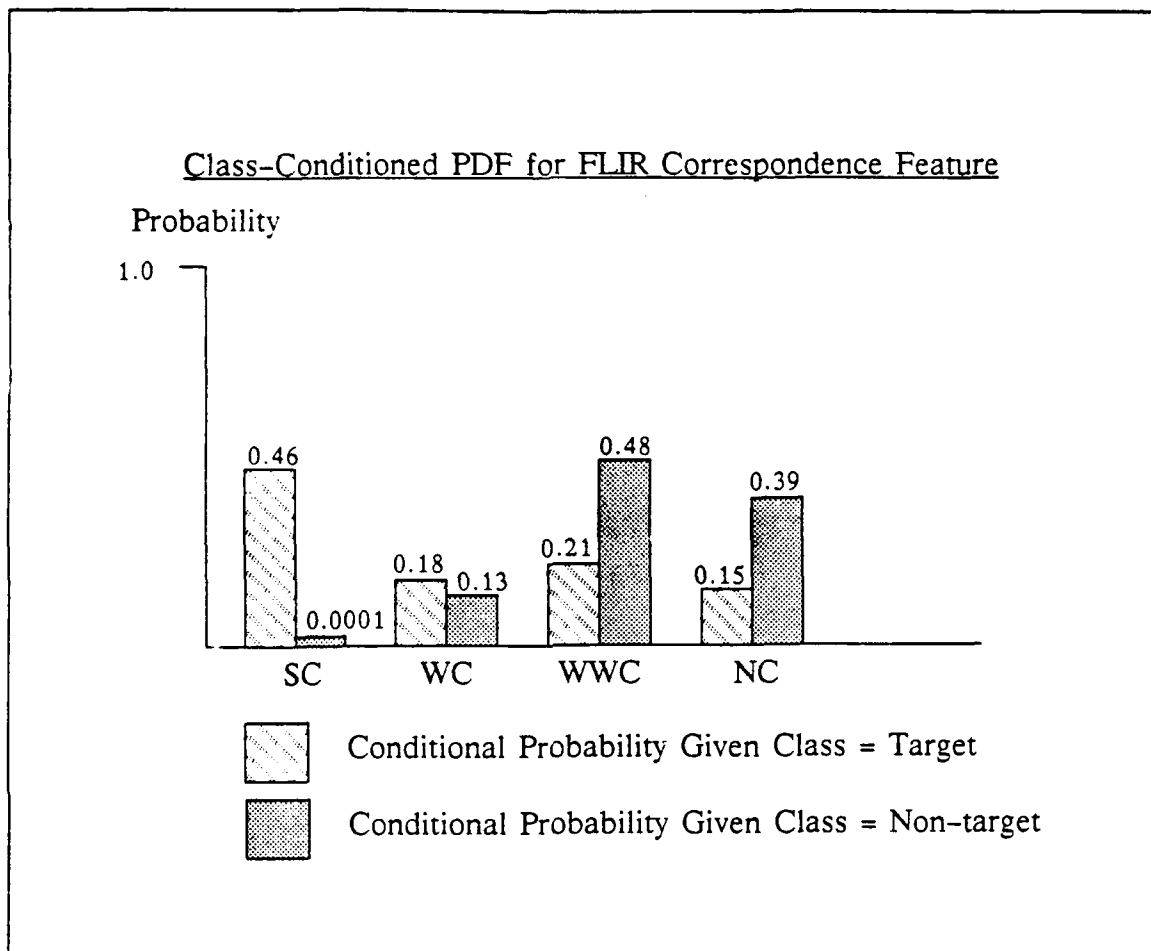


Figure (5-2). Discrete class-conditioned PDF for FLIR image correspondence feature.

corresponds to an impossible event. The possibility of a FLIR non-target region with a SC was viewed as unlikely, rather than impossible. Thus, the value was set as shown in Figure (5-2).

### 5.5 Conclusions

A set of single sensor features was selected for further study from a larger set of features using the best feature approach, based on the criterion of minimizing single feature probability of error. Three FLIR image features and three range image features were selected using this technique. The performance of the various combinations of



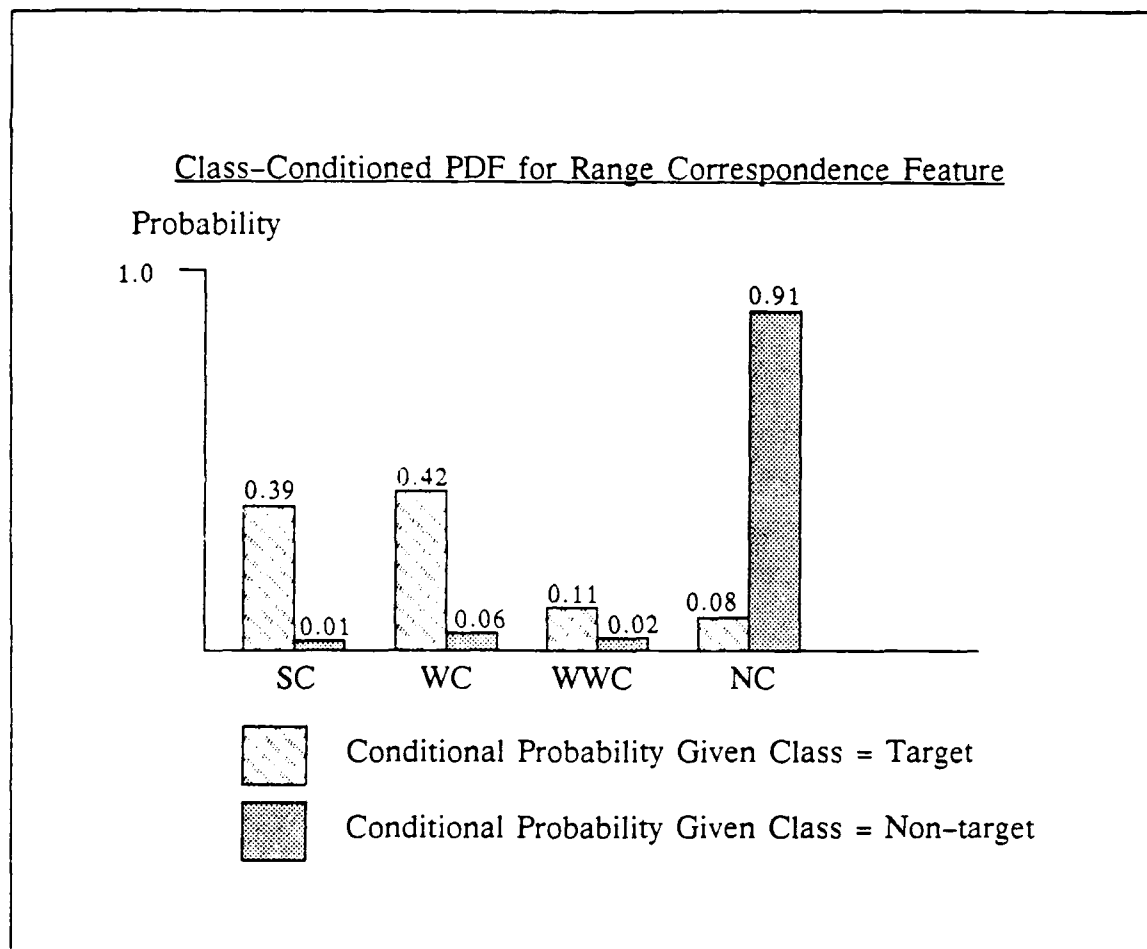


Figure (5-3). Discrete class-conditioned PDF for range image correspondence feature.

these features is discussed in Chapter VI. Based on the good performance obtained with the features selected in this manner, it was concluded that the best features approach to choosing features was adequate for the present problem.

The multiple sensor correspondence feature was developed to exploit the observation that targets lie in the same space, regardless of which sensor viewed the scene, while segmented non-target regions do not tend to behave in this manner. The correspondence feature took four mutually exclusive values. Each of the values provided useful information about the properties of a region in the non-dominant sensor image which was cued

by a region in the dominant sensor image.

Allowances were made for small errors in the locations of the common pixels in the various percentages used to assign correspondence feature values. More accurate and consistent geometric registration would allow these percentages to be raised. One potential result of raising these tolerances would be that fewer non-target regions would be assigned correspondence feature values of SC and WC. This would, in turn, would result in the correspondence feature being a better discriminator of targets and non-targets.

Figures (5-2) and (5-3) show that the correspondence feature for range images is a better discriminant of targets and non-targets than the correspondence feature for FLIR images. The major factor contributing to this result is higher noise in the range imagery. High noise on a range image target contributes directly to high absolute errors associated with a plane fit to that region, often to the extent that even the relaxed error threshold will not pass a target region as being planar. Thus, proportionally more FLIR target regions acquired a correspondence feature value of NC than range image target regions. The large role of heuristics in the range image segmentation process also contributed to this result. Specifically, while the planarity test was found to be an excellent selector of target pixels, it often allowed a larger fraction of the scene to pass the initial segmentation test than the brightness threshold on the FLIR image. Heuristics were used to reject most of the non-target pixels. However, the result was that segmented non-targets in FLIR images were more likely to acquire correspondence feature values of WWC than segmented non-targets in range images. A better criterion for WWC for segmented FLIR regions, perhaps including surface orientation information, would probably improve the performance of the FLIR correspondence feature.

The correspondence feature is a novel feature which, as is discussed in Chapter VI, provided very useful information to the multiple sensor class estimation process. The correspondence feature may only be obtained in a multiple sensor system. Performance improvements resulting from incorporation of this information into the decision process

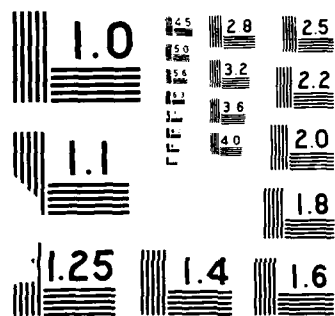
AD-A207 577

MULTIPLE SENSOR FUSION FOR DETECTING TARGETS IN FLIR  
(FORWARD-LOOKING INF. (U) AIR FORCE INST OF TECH  
WRIGHT-PATTERSON AFB OH SCHOOL OF ENGI. M C ROGGENANN  
MAY 89 AFIT/DS/ENP/89-1 F/G 17/5.1

2/2

UNCLASSIFIED

140  
141  
142



advocate strongly for use of multiple sensor target detection systems.

## VI. Single Sensor and Multiple Sensor Target Detection

### 6.0 Introduction

The problem discussed in this chapter is that of automatically estimating the class, target or non-target, of segmented regions in FLIR and range images. Topics discussed include formulation of the Bayesian decision problem, single sensor and multiple sensor target detection algorithms, and the performance obtained with each approach. Single sensor and multiple sensor target detection approaches were distinguished by the use of correspondence feature information in the multiple sensor cases. Use of correspondence feature information was found to improve target detection rates in every case, while reducing, or not affecting the false alarm rates.

Two single sensor target detection algorithms were developed: (1) FLIR-only; and (2) range-only. These algorithms estimated the class of segmented regions based on information available from only one sensor. An exhaustive search of the three best features for each sensor, which were discussed in Chapter V, was conducted to obtain optimal performance from the single sensor cases. The single sensor cases provided a baseline performance for comparison to the multiple sensor cases.

Three multiple sensor target detection approaches were examined: (1) FLIR looking into range (FLIR/range), where the class of segmented regions in FLIR images was estimated using feature information obtained from the single sensor FLIR image features and from the correspondence feature; (2) range looking into FLIR (range/FLIR), the reciprocal of (1), where the class of segmented regions in range images was estimated using single sensor range image feature information and correspondence feature information; and (3) the single decision (SD) algorithm, in which the joint spatial occupancy of segmented regions in space was resolved, and a single decision made for each segmented region of space, regardless of which sensor the region appeared in. All of the multiple

sensor approaches used correspondence feature information in the class estimation process. The SD algorithm also used the correspondence tables computed during the correspondence feature measurement (see Chapter V) to resolve joint spatial occupancy issues between segmented regions in the images. An exhaustive search of the three best single sensor features was conducted to determine the best set of features to use in conjunction with the correspondence feature.

In the FLIR/range and range/FLIR algorithms a single class estimate was computed for each segmented region in the dominant sensor image. The concept of dominant and non-dominant sensor images was defined in Chapter V. To reiterate, the dominant sensor image was the sensor image used to drive the search of cued regions in the other, non-dominant sensor image during correspondence feature measurement. For the FLIR/range algorithm, the FLIR image was the dominant sensor image. For the range/FLIR algorithm, the range image was the dominant sensor image. These algorithms explored the concept of using the non-dominant sensor to assist the dominant sensor. It is important to note that the upper bound on the number of target detection opportunities for the FLIR/range and range/FLIR algorithms was the set of targets segmented in the dominant sensor image. This was not the case for the SD algorithm.

The SD algorithm contained a rule for determining when segmented regions in both sensor images occupied the same space. When joint spatial occupancy was detected, a single class estimate was made for that region of space. In addition, the SD algorithm determined where regions of space were segmented by only one sensor image, and also made a class estimate for those regions. Thus, the upper bound on the number of target detection opportunities for the SD algorithm was the the union of the sets of targets segmented in each sensor image. This set is always at least as large as the set of targets segmented in one of the sets of sensor images. In the present data base, the set of target opportunities for the SD algorithm was larger than the set of target opportunities for any of the other detection algorithms.

Feature values and image truth, in the form of a target or non-target label for each segmented region of each sensor image, were obtained for all segmented regions in the data base. Image truth and feature values were stored indexed to the image file name and the region label using the region labeling scheme described in Chapter V. This information was stored and accessed as needed.

The remainder of this chapter is organized as follows. Background information pertinent to the approach taken is presented in the next section. The mathematical formulation of the Bayesian class estimation problem is then discussed. Training, testing, and performance measures are presented. This is followed by a discussion of the implementation of the single and multiple sensor target detection algorithms. Next, image truth and data base considerations are discussed. Selection of optimum feature sets for each detection system, and performance results are then discussed. Conclusions are drawn in the final section.

## **6.1 Background and Approach**

The target detection problem was approached as a two-class estimation problem in which the class estimate was based on a single temporal observation (Melsa and Cohn, 1978:21-53). Bayesian decision theory was adopted to perform the class estimation process. In particular, the Bayesian minimum probability of error decision rule (Melsa and Cohn, 1978:42; Devijver and Kittler, 1982:33-43), also known as the Maximum a Posteriori (MAP) decision criterion, was used.

Use of the MAP decision rule in conjunction with a single feature requires that the class-conditioned probabilities for the feature value be exhibited. The feature values obtained from the data base constituted finite samples of inherently continuous random variables. The parametric form of the density functions governing these random variables, if such density functions exist, were unknown, a common problem in pattern recognition (Fukunaga, 1972:165; Devijver and Kittler, 1982:63). To address this problem, discrete class-conditioned probability density functions (PDF) for each feature were



measured using a histogram approach described in Chapter V, and discussed in detail later in this chapter.

When multiple features are used in the class estimation process, as was the case here, the class-conditioned probabilities for the feature set is required. This quantity is a joint conditional probability of dimension equal to the number of features used. Obtaining a reasonable estimate for multi-dimensional conditional probabilities is, in many cases of interest, quite difficult (Duda et al, 1979b:83; Cheeseman, 1983:199; Cheeseman, 1985:1003-1004). To alleviate the difficulty associated with obtaining multi-dimensional class-conditioned probabilities, the features were assumed to be conditionally independent.

The assumption of conditional independence simplified the problem of exhibiting the required multi-dimensional class-conditioned probabilities. While this assumption is often, in practice, less than perfectly realized, this assumption has precedent, and has been found useful for similar problems (Lewis, 1962; Duda et al, 1976:1080; Duda et al, 1979b:83-84; Cheeseman, 1985:1004). The major concern with assuming conditional independence for data which is not conditionally independent is that single feature performance measures cannot be readily extrapolated to a prediction of multiple feature performance (Lewis, 1962:173). Unexpected performance results can arise due to unaccounted for dependences between the features (Duda et al, 1979b:88-92; Cheeseman, 1983:198). However, the assumption of conditional independence was found to be useful for the present work.

Prior densities for the classes must also be known, or assumed, to use the MAP decision criterion. The prior densities were observable over the database. However, the values obtained were functions of both the background environment and the target density provided during the data collection. Since there was no reason to suppose these densities would be equivalent to the observed values under different data collection conditions, the prior densities were set to be equally likely. The assumption of equally likely

prior densities in the absence of good reason to choose otherwise has been called the 'principle of indifference' (Cheeseman, 1985).

Training and testing are always critical issues in classifier design and evaluation. The goal of training and testing was to obtain a sample-based estimate of the actual error rate of the classifier which would be observed by testing on a large amount of equivalently distributed data. Training of the classifier consisted of exhibiting the class-conditioned PDFs of the features for a subset of the entire database. Testing was accomplished by tabulating the performance of the class estimation algorithm on a subset of the database disjoint from the training subset.

Several methods of selecting training and testing subsets have been developed (Devijver and Kittler, 1982:343-359). The hold-one-out method was adopted for this project (Foley, 1972:618; Devijver and Kittler, 1982:356-357). If  $Q$  samples are available, in this method one sample is withheld while the classifier is trained on the remaining  $(Q-1)$  samples. The classifier is then tested on the withheld sample, and the results are tabulated. This procedure is repeated  $Q$  times, with a different sample withheld each time. Results of such an exercise constitute the average performance across the  $Q$  samples. In the present case, images constituted the samples even though the images, in general, contained more than one segmented region. Given a finite number of samples, this method is the preferred method of obtaining an estimate of the error rate if sufficient computational resources are available.

Preference for the hold-one-out method stems from its highly efficient use of the available data, and from the fact that the estimate of the error rate obtained using this method is approximately unbiased, regardless of the underlying distributions of the features (Devijver and Kittler, 1982:356). An estimator of a statistical quantity, in this case the sample-based estimate of the actual error rate, is said to be unbiased if the expected value of the estimator equals the value of the parameter being estimated (Keeping, 1962:101). Other methods of estimating the error rate, such as the resubstitution

method, the hold-out method, and the rotation method, are unduly biased (that is, they give overly optimistic or pessimistic estimates of the error rate), or use the information available in the database less efficiently (Devijver and Kittler:1982:353-359).

Comparison of competing designs is important to evaluating the performance of class estimation algorithms. An error counting method for obtaining the 95% confidence interval on the error rate was adopted (Devijver and Kittler, 1982:346-349). The 95% confidence interval provides a range of values for the error rate within which the actual error rate for an infinite amount of equivalently distributed data would lie with 95% probability. Two additional performance measures were also used: (1) the target detection rate; and (2) the rate of false alarms per detection declaration. The total error rate addresses all classification errors, while the target detection rate and the rate of false alarms per detection declaration isolate the two types of possible classification errors.

## 6.2 Formulation of the Bayesian Class Estimation Problem

The MAP decision criterion (Melsa and Cohn, 1978:38-44) was used to estimate the class of segmented regions. The specific problem was to estimate to which of the classes,  $\{\theta_k\}$ , each segmented region belonged. Information available to the class estimation process consisted of a set of feature measurements,  $\{f_i(j)\}$ , the class-conditioned probabilities of observing  $\{f_i(j)\}$ ,  $p(\{f_i(j)\}|\theta_k)$ , and estimates of the probabilities of observing the classes,  $p(\theta_k)$ , called the prior densities. The notation  $f_i(j)$  refers to the measurement of the  $i^{th}$  feature in the  $j^{th}$  bin (see Chapter V). Computation of the single feature class-conditioned probabilities,  $p(f_i(j)|\theta_k)$ , is described in the next section. The set  $\{\theta_k\}$  consisted of two classes,  $\theta_1 = \text{target}$ , and  $\theta_2 = \text{non-target}$ . For reasons discussed previously, the prior densities were set equally likely:  $p(\theta_1) = p(\theta_2) = 0.5$ .

When multiple features are used, the class-conditioned probabilities of interest are the  $p(\{f_i(j)\}|\theta_k)$ . The features were assumed to be conditionally independent, so that:

$$p(\{f_i(j)\}|\theta_k) = \prod_{i=1}^I p(f_i(j)|\theta_k) \quad (6-1)$$

where  $I$  is the number of features.

For the present problem, the MAP criterion may be stated as: given an observation ,  $\{f_i(j)\}$ , choose the most likely class,  $\theta_k$  (Melsa and Cohn, 1978:42). The probability of the occurrence of the class  $\theta_k$  given the observation  $\{f_i(j)\}$  must be computed. Bayes rule provides for this computation:

$$p(\theta_k | \{f_i(j)\}) = \frac{p(\{f_i(j)\} | \theta_k) p(\theta_k)}{p(\{f_i(j)\})} \quad (6-2)$$

In Equation (6-2) the probability  $p(\{f_i(j)\})$  is the probability of observing the feature set  $\{f_i(j)\}$ , given by:

$$p(\{f_i(j)\}) = \sum_{k=1}^K p(\{f_i(j)\} | \theta_k) p(\theta_k) \quad (6-3)$$

Thus, the class estimation problem was reduced to the problem of computing  $p(\theta_1 | \{f_i(j)\})$  and  $p(\theta_2 | \{f_i(j)\})$ . When  $p(\theta_1 | \{f_i(j)\}) > p(\theta_2 | \{f_i(j)\})$  then the class estimate for the region was  $\theta_1$ , otherwise the region was estimated as being a member of class  $\theta_2$ .

The underlying mathematical principle of this class estimation technique was quite simple. In general, the major problem encountered with using this approach is obtaining reasonable estimates for the class-conditioned probabilities and the prior densities (Duda et al, 1979b:83; Garvey and Lowrance, 1981:3-5; Devijver and Kittler, 1982:62-63; Lowrance and Garvey, 1983:2-9; Cheeseman, 1983:198; Cheeseman, 1985). The ultimate justification for any approach to exhibiting the required conditional and prior probabilities lies in the performance obtained, which was judged to be quite good.

### 6.3 Training, Testing, and Performance Measures

Training the classifier consisted of obtaining estimates for the discrete class-conditioned PDFs,  $p(f_i(j) | \theta_k)$ . These estimates were obtained from a subset of the data base called the training set. Testing of the classifier was accomplished by measuring the

performance of the classifier on the subset of the data base not included in the training set, called the test set. Three measurements of performance were used to evaluate the various classification algorithms: (1) detection rate,  $P_d$ ; (2) the rate of false alarms per detection declaration,  $FAR$ ; and (3) the total error rate,  $P_e(tot)$ . A false alarm was defined as a non-target region which was incorrectly classified as a target.

Training and testing subsets of the data base were selected using the hold-one-out method. This technique was implemented by withholding a single matched pair of FLIR and range images from the data base, and using the remainder of the data base for training. The performance of the classifier was measured on the withheld samples, and the process was repeated until all the matched pairs of images had been withheld once. During this process, performance on the test samples was continuously tabulated.

The discrete class-conditioned PDFs of the features were measured on the training set using a histogram approach with equally spaced bins, as discussed in Chapter V. Let  $N_b$  be the number of bins, and let  $f_i(max)$  and  $f_i(min)$  be the maximum and minimum excursions of the  $i^{th}$  feature observed in the training set. The histogram approach to exhibiting the required PDFs was implemented by dividing the interval  $[f_i(min), f_i(max)]$  into  $N_b$  equally spaced bins of width,  $W_b$ :

$$W_b = \frac{f_i(max) - f_i(min)}{N_b} \quad (6-4)$$

A feature value,  $f_i$ , fell in the  $j^{th}$  bin when:

$$(f_i(min) + (j-1)W_b) \leq f_i < (f_i(min) + jW_b) \quad (6-5)$$

where the bins were indexed by  $j$ ,  $1 \leq j \leq N_b$ . The bins were dimensionless, and hence, the notation  $f_i(j)$  to denote the occurrence of the  $i^{th}$  feature for a region having a value in the  $j^{th}$  bin. The number of occurrences of the event  $\{class=\theta_k \text{ and } f_i(j)\}$ ,  $n(f_i(j) \text{ and } \theta_k)$ , in the training set were counted. The number of occurrences of each class,  $n(\theta_k)$  in the training set were also counted. The class-conditioned probabilities of observing the  $i^{th}$  feature in the  $j^{th}$  bin,  $p(f_i(j) | \theta_k)$ , were then computed using Equation

(5-1). The collection of these probabilities for both classes, for all bins constituted the discrete class-conditioned PDF for a feature. Empirically obtained values for  $N_b$  were used:  $N_b = 15$  for FLIR data, and  $N_b = 7$  for range data.

When testing was performed, there was no guarantee that the feature values observed,  $f_i$ , would fall in the interval  $[f_i(\min), f_i(\max)]$ , since the test data was not included in the training set. In the implementation,  $f_i$  which were greater than  $f_i(\max)$  were mapped to the bin  $N_b$ , and  $f_i$  which were less than  $f_i(\min)$  were mapped to bin 1.

As testing was conducted, four performance-related variables were tabulated: (1) the number of target opportunities,  $N_t$ ; (2) the number of targets correctly classified,  $N_t(\text{corr})$ ; (3) the number of non-target opportunities,  $N_{nt}$ ; and (4) the number of non-targets correctly classified,  $N_{nt}(\text{corr})$ . Three performance measures were computed: (1) the target detection rate,  $P_d$ :

$$P_d = \frac{N_t(\text{corr})}{N_t} \quad (6-6)$$

(2) the rate of false alarms per detection declaration,  $FAR$ :

$$FAR = \frac{(N_{nt} - N_{nt}(\text{corr}))}{(N_{nt} - N_{nt}(\text{corr})) + N_t(\text{corr})} \quad (6-7)$$

and, (3) the total error rate,  $P_e(\text{tot})$ :

$$P_e(\text{tot}) = \frac{(N_t - N_t(\text{corr})) + (N_{nt} - N_{nt}(\text{corr}))}{N_t + N_{nt}} \quad (6-8)$$

The concept of the 95% confidence interval (Keeping, 1962:96-101; Devijver and Kittler, 1982:346-349) was used to compare  $P_e(\text{tot})$  for the various detection algorithms. The sample-based estimate of the variance of the total error rate,  $\sigma_e^2(\text{tot})$  is given by (Devijver and Kittler, 1982:347):

$$\sigma_e^2(\text{tot}) = \frac{P_e(\text{tot})(1 - P_e(\text{tot}))}{N_{\text{tot}}} \quad (6-9)$$

where  $N_{\text{tot}}$  is the total number of samples tested. It is possible to show that, with 95%

probability, the value of  $P_e(tot)$  that would be observed for a large amount of equivalently distributed data lies in the interval  $[(P_e(tot) - 1.96\sigma_e(tot)), (P_e(tot) + 1.96\sigma_e(tot))]$  (Devijver and Kittler, 1982:347-349). This interval is called the 95% confidence interval for  $P_e(tot)$ .

#### 6.4 Detection Algorithm Implementations

The FLIR-only and range-only detection algorithms used feature information available from only one sensor to make class estimates. These algorithms were implemented by computing a class estimate for each segmented region in the images using single sensor feature information and the training and testing techniques discussed above.

The multiple sensor algorithms made class estimates using multiple sensor information for segmented regions which were viewed completely by both sensor images. The FLIR images viewed all of the segmented regions in the range images, but the converse was not true. Many FLIR targets were outside the field of view of the range sensor, and many segmented non-target regions in the FLIR imagery were only partially viewed by the associated range image. Segmented regions not viewed completely by both sensor images were not considered by the multiple sensor algorithms.

The FLIR/range and range/FLIR detection algorithms computed class estimates for segmented regions in the dominant sensor image. Single sensor feature information from the dominant sensor image, and the multiple sensor correspondence feature were used in these class estimation processes. Class estimation, training, and testing were accomplished using the techniques outlined above.

The SD detection algorithm computed a class estimate for each segmented region of space viewed completely by both sensor images, regardless of which sensor image the segmented regions appeared in. Single sensor feature information, correspondence feature information, and, under certain conditions, feature information from both sensor images was used to make class estimates. Class estimation, training, and testing were

conducted using the methods described above.

In the SD algorithm, both sensor images were used sequentially as the dominant sensor image to measure the correspondence feature and to make the appropriate entries in the correspondence tables for all segmented regions. The probabilities  $p(\theta_k | \{f_i(j)\})$  were then computed for all regions using single sensor features and the correspondence feature, and under certain conditions, feature information from both sensors. An exhaustive search of the correspondence tables was then conducted to resolve the joint spatial occupancy issues, allowing only one class estimate to be made for each segmented region of space. The process of resolving the joint spatial occupancy issues was called deconfliction.

The deconfliction rule performed an exhaustive search of the correspondence tables for FLIR and range images. The images were searched sequentially, with the FLIR image arbitrarily selected as the first image searched. The correspondence feature tables of interest to the deconfliction algorithm were the strong correspondence table,  $SC(m)$ , and the weak correspondence table,  $WC(m)$ , since these correspondence feature values indicated that segmented regions in both sensor images jointly occupied the same space.

Correspondence table entries were defined in Chapter V to be of the form  $SC(m) = n$  and  $WC(m) = n$ , which was interpreted as: "the  $m^{th}$  region in the dominant sensor image has a SC or WC, appropriately, with the  $n^{th}$  region in the non-dominant sensor image". This notation is now refined with a subscript,  $F$  or  $R$ , to indicate that FLIR or range, respectively, was the dominant sensor image. Thus,  $SC_F(m) = n$  implies that the  $m^{th}$  region in the FLIR image had a SC with the  $n^{th}$  region in the range image.

Two special cases of joint spatial occupancy were of interest: (1) mutual correspondence; and (2) non-mutual correspondence. Mutual correspondences occurred under the following cases of correspondence table entries:

- (1)  $SC_F(m) = n$ , and  $SC_R(n) = m$
- (2)  $SC_F(m) = n$ , and  $WC_R(n) = m$



$$(3) WC_F(m) = n, \text{ and } SC_R(n) = m$$

$$(4) WC_F(m) = n, \text{ and } WC_R(n) = m$$

Mutual correspondence occurred when segmented regions occupied the same space to a good approximation. For example, two well segmented targets would typically have a mutual SC. Non-mutual correspondences occurred under all other combinations of the SC and WC tables, specifically:

$$(1) SC_F(m) = n, \text{ but } SC_R(n) \neq m \text{ and } WC_R(n) \neq m$$

$$(2) WC_F(m) = n, \text{ but } SC_R(n) \neq m \text{ and } WC_R(n) \neq m$$

$$(3) SC_R(n) = m, \text{ but } SC_F(m) \neq n \text{ and } WC_R(m) \neq n$$

$$(4) WC_R(n) = m, \text{ but } SC_F(m) \neq n \text{ and } WC_R(m) \neq n$$

Non-mutual correspondence occurred when more than one region in one image corresponded with a single region in the other image.

For example, a non-mutual correspondence occurred when a tank and a jeep segmented distinctly in the range image, but were segmented as a single region in the corresponding FLIR image. In this case the range image tank had either a SC or a WC with the FLIR region, the range image jeep had a WC with the FLIR region, and the FLIR image region had either a WC or a SC with the range image tank. Thus, the range image tank and the FLIR image region had a mutual correspondence, while the range image jeep and the FLIR image region had a non-mutual correspondence. The spatial deconfliction algorithm contained a rule for resolving such occurrences.

Given the above partitioning of spatial correspondences, three possible cases of joint spatial occupancy confronted the SD algorithm: (1) mutual correspondence; (2) non-mutual correspondence; and (3) no spatial correspondence between segmented regions. Multiple sensor information fusion and spatial deconfliction were handled in the following manner:

(1) Mutual correspondence: Multiple sensor feature information was merged by assuming conditional independence, and computing new class-conditioned probabilities for the

combined feature set using:

$$p_{new}(\{f_i(j)\}_F \cup \{f_i(j)\}_R | \theta_k) = p(\{f_i(j)\}_F | \theta_k) p(\{f_i(j)\}_R | \theta_k) \quad (6-10)$$

where the subscripts  $F$  and  $R$ , for FLIR and range, respectively, indicate which sensor image provided the feature set. New estimates of  $p(\theta_k | \{f_i(j)\})$  were computed using Equation (6-10) in Equation (6-2), and the MAP criterion was applied to obtain a new class estimate. A flag was raised in a table associated with the indices of the appropriate regions to insure that the regions were never reconsidered by the SD algorithm.

(2) Non-mutual correspondence: The measure of confidence:

$$\Delta = |p(\theta_1 | \{f_i(j)\}) - p(\theta_2 | \{f_i(j)\})| \quad (6-11)$$

was computed for both regions, and the region with the largest  $\Delta$  was used to make the class estimate. A flag was raised in a table associated with the indices of the regions considered to insure the regions were never reconsidered by the SD algorithm.

(3) No correspondence: This occurred when a region had a correspondence feature value of WWC or NC. A check of all regions in the other image was made to see if any SC or WC existed to the region in question. If a SC or a WC was found to a region possessing a WWC or a NC, and if the region possessing the SC or WC had not already been used in (1) or (2) above, then the region possessing the SC or WC, was used to make the class estimate for that region of space. Otherwise, the class estimate was made using the available information for the region possessing the WWC or NC. A flag was raised in a table associated with the indices of the regions considered to insure the regions were never reconsidered by SD algorithm.

The deconfliction rule allowed the class of segmented regions of space to be estimated without redundancy. Other deconfliction rules are possible. For example, if high quality estimates of the relative positions and pointing angles of the sensors had been available, registration of the regions in an (azimuth angle, elevation angle) space

would have been possible.

## **6.5 Image Truth and Data Base Considerations**

Image truth was obtained through manual inspection of the segmented images in the data base. Segmented regions were labeled with an integer value using the region labeling scheme discussed in Chapter V. Labeled images were displayed in conjunction with the associated region label, and a target/non-target determination was made for each region. The results were recorded and stored for easy access.

A segmented region was labeled as a target if it contained a target, more than one target, or a subjectively evaluated 'significant' portion of a target. All other regions were labeled non-targets.

The FLIR image data base contained 97 images. The FLIR data base contained 230 segmented target regions, 153 of which were viewed completely by the corresponding range images. It also contained 320 segmented non-target regions, of which 23 were completely viewed by the associated range images.

The number of segmented target regions is different than the number of segmented targets reported in Chapter III because the segmentation scoring method and the target region counting method were different. Successful segmentations were scored if targets visible in an image appeared in the segmented version of the image. However, in several cases targets parked very close to each other were segmented as a single region. For example, when a truck was occluding a jeep the result was that, quite often, the tank and the jeep were segmented as a single region. This situation was scored as two successful segmentations for two opportunities to segment a target. However, only one target region appeared in the image, which contained both vehicles.

The range image data base consisted of 57 images containing 121 targets and 276 non-target regions. Because some range image corresponded to more than one FLIR image, the total number of range image targets viewed by running through all 97 FLIR

images was 207, and the number of non-target regions viewed was 463.

The disparity between the number of target regions in the range image data base, 207, and the number of target regions in the FLIR data base, 153, was a result of three anomalies in the data bases. First, multiple FLIR targets were occasionally segmented as a single region, where this was never observed to occur for range images, having the effect of decreasing the number of FLIR target regions relative to the number of range target regions. Second, targets in the range image data base were more likely to be fractured into two pieces by segmentation due to noise, having the effect of increasing the number of range target regions relative to the number of FLIR target regions.

The third anomaly between the FLIR and range data bases involved a group of targets at approximately 860 m range which appeared in the foreground of several images of targets at approximately 1700 m range. In the FLIR image data base large portions of these target were merged with the background due to the gain and brightness settings of the FLIR being adjusted to view the targets at 1700 m. (In fact, these targets were not as 'visible' in the FLIR segmentation scoring.) The targets at 860 m were typically 'chopped up' by the brightness threshold, and discarded by the heuristics. The range sensor had no adjustments analogous to the gain and brightness settings of a FLIR, and the targets at 860 m were, quite frequently, segmented accurately in the range images. This also had the effect of increasing the number of target regions in the range data base relative to the FLIR data base.

The deconfliction algorithm, described in the previous section, was used to determine the number of segmented target and non-target regions in the space viewed by both sensors. The result was that 217 target regions and 484 non-target regions were found. Included in the 217 target regions were 36 target regions not segmented in the FLIR image data base, and 11 target regions not segmented in the range image data base.

## 6.6 Optimum Feature Sets and Performance

The three best features for each type of sensor image, discussed in Chapter V, were examined exhaustively to obtain the optimal performance for each detection algorithm. Optimum performance for three performance measures was obtained: (1) minimum total error rate,  $P_e(tot)$ ; (2) maximum detection rate,  $P_d$ ; and (3) minimum rate of false alarms per detection declaration,  $FAR$ .

From Chapter V, the three best FLIR features were: (1) complexity; (2) length-to-width ratio; and (3) contrast of the means. The three best range image features were: (1) length-to-width ratio; (2) the absolute difference of the standard deviations; and (3) complexity. These features will be referred to by sensor and index in the discussion which follows. For example, FLIR:2 refers to the FLIR length-to-width ratio feature.

The features found to give optimal performance for each measure are listed in Tables (6-1), (6-2), and (6-3). Table (6-1) lists the best features for minimum  $P_e(tot)$ . Table (6-2) lists the best features for maximum  $P_d$ . Table (6-3) lists the best features for minimum  $FAR$ .

| Table (6-1). Features giving minimum $P_e(tot)$ . |                     |
|---|---------------------|
| Algorithm   | Feature Index       |
| FLIR  | FLIR: 1,2,3         |
| Range   | Range: 1,2,3        |
| FLIR/Range  | FLIR: 1,3           |
| Range/FLIR  | Range: 1            |
| SD  | FLIR: 1,3; Range: 1 |

The values obtained for  $P_e(tot)$ , the 95% confidence interval on  $P_e(tot)$ ,  $P_d$ , and  $FAR$  for each algorithm as a function of the performance measure optimized are displayed in Tables (6-4), (6-5), and (6-6). Table (6-4) shows performance for the minimum  $P_e(tot)$  criterion. Table (6-5) gives performance for the maximum  $P_d$

| Table (6-2). Features giving maximum $P_d$      |   |
|---|---|
| Algorithm                                       | Feature Index   |
| FLIR<br>Range<br>FLIR/Range<br>Range/FLIR<br>SD | FLIR: 1,2,3<br>Range: 1,3<br>FLIR: 1,3<br>Range: 1<br>FLIR: 1,3; Range: 1 |

criterion. Table (6-6) gives performance for the minimum  $FAR$  criterion. Tables of absolute performance for every combination of the features is provided in Appendix C.

| Table (6-3). Features giving minimum $FAR$      |   |
|---|---|
| Algorithm                                       | Feature Index   |
| FLIR<br>Range<br>FLIR/Range<br>Range/FLIR<br>SD | FLIR: 1,2,3<br>Range: 1,2,3<br>FLIR: 1,3<br>Range: 1,2,3<br>FLIR: 1,2,3; Range: 1,2,3 |

| Table (6-4). Performance achieved with minimum $P_e(tot)$ . |            |                         |       |       |
|---|------------|-------------------------|-------|-------|
| Algorithm   | $P_e(tot)$ | 95% Confidence Interval | $P_d$ | $FAR$ |
| FLIR  | 0.131      | (0.081,0.181)           | 0.856 | 0.008 |
| Range   | 0.136      | (0.110,0.162)           | 0.681 | 0.151 |
| FLIR/Range  | 0.091      | (0.048,0.133)           | 0.902 | 0.007 |
| Range/FLIR  | 0.060      | (0.042,0.078)           | 0.952 | 0.132 |
| SD  | 0.069      | (0.050,0.087)           | 0.926 | 0.137 |

The power of the SD algorithm to detect targets not segmented in both images can

be seen by noting that for the cases show in Tables (6-4) and (6-5) the SD algorithm correctly detected 6 of 11 targets not segmented in the range image data base, and 26 of 36 targets not segmented in the FLIR image data base. For the case displayed in Table (6-6) the SD algorithm correctly detected 5 of 11 targets not segmented in the range image data base and 28 of 36 targets not segmented in the FLIR image data base.

| Table (6-5). Performance achieved with maximum $P_d$ . |            |                         |       |       |
|--|------------|-------------------------|-------|-------|
| Algorithm  | $P_e(tot)$ | 95% Confidence Interval | $P_d$ | $FAR$ |
| FLIR   | 0.131      | (0.081,0.181)           | 0.856 | 0.008 |
| Range  | 0.179      | (0.150,0.208)           | 0.850 | 0.336 |
| FLIR/Range   | 0.091      | (0.048,0.133)           | 0.902 | 0.007 |
| Range/FLIR   | 0.060      | (0.042,0.078)           | 0.952 | 0.132 |
| SD   | 0.069      | (0.050,0.087)           | 0.926 | 0.137 |

Examination of Tables (6-4), (6-5) and (6-6) show clearly that use of multiple sensor information improves target detection performance by every measure used. The statistical significance of the performance improvement is best explained by examining the 95% confidence intervals arising from minimizing  $P_e(tot)$ . When the intersection of the confidence intervals for two competing algorithms is empty or small, then it can be claimed with high confidence that the algorithm with lower  $P_e(tot)$  represents a significant improvement over the other algorithm. Thus, Table (6-4) shows that the range/FLIR and SD algorithms are significantly better, in the minimum  $P_e(tot)$  sense, than either single sensor algorithm.

The case for the FLIR/range algorithm being a significant improvement over the single sensor cases is somewhat weaker due to the large overlap of the confidence intervals. Thus, it cannot be stated with high confidence that the FLIR/range algorithm is significantly better than the single sensor approaches, even though gratifying improvements in both  $P_e(tot)$  and  $P_d$  were obtained. Failure of the FLIR/range algorithm to

| Table (6-6). Performance achieved with minimum <i>FAR</i> . |            |                         |       |            |
|---|------------|-------------------------|-------|------------|
| Algorithm   | $P_e(tot)$ | 95% Confidence Interval | $P_d$ | <i>FAR</i> |
| FLIR  | 0.131      | (0.081,0.181)           | 0.856 | 0.008      |
| Range   | 0.136      | (0.110,0.162)           | 0.681 | 0.151      |
| FLIR/Range  | 0.091      | (0.048,0.133)           | 0.902 | 0.007      |
| Range/FLIR  | 0.061      | (0.043,0.079)           | 0.903 | 0.101      |
| SD  | 0.076      | (0.056,0.096)           | 0.872 | 0.110      |

meet this measure of statistical significance is a direct consequence of the relatively small number of samples in the FLIR data base (176) compared to the number of samples in the range data base (670) and the SD (700) data base.

A word of caution is required for the FLIR and FLIR/range *FAR*. The FLIR data base contained a very small number of non-target regions viewed completely by both sensors (23). The *FAR*s show in Tables (6-4), (6-5), and (6-6) represent misclassification of one non-target region in every case.

The FLIR and FLIR/Range *FAR* performance can be extrapolated by assuming that in a more reasonable data set the ratio of segmented non-targets to segmented targets would remain constant independent of which subset of the field of view was used ( that is,  $320/230 = 1.391$ ), as would the rate of misclassification of non-target regions (that is,  $1/23 = 0.043$ ). Then for the 153 segmented target regions there would be  $153 \times 1.391 \approx 213$  segmented non-targets of which  $0.043 \times 213 \approx 9$  would be misclassified. Using these new figure, the entries in Table (6-4) for the FLIR algorithm would become:  $P_3(tot) = 0.085$ , 95% confidence interval = (0.056,0.114),  $P_d = 0.856$ , and *FAR* = 0.064. The entries in Table (6-5) for the FLIR/range algorithm would become:  $P_e(tot) = 0.066$ , 95% confidence interval = (0.041,0.091),  $P_d = 0.902$ , and *FAR* = 0.061. Even with this extrapolation it is not clear that the FLIR/range algorithm is a significant improvement over the FLIR-only algorithm. However, it is possible that a more reasonable estimate of



the FLIR-related *FAR* has been obtained.

## 6.7 Conclusions

Numerous simplifications were made to achieve the performance obtained. In particular, the assumption of conditional independence between all features, the use of a histogram approach to estimating class-conditioned probabilities, and use of the suboptimal best features approach to selecting a feature set were potentially risky assumptions. The ultimate justification for the utility of these assumptions lies in the performance obtained. It was concluded, based on the performance exhibited in Tables (6-4), (6-5), and (6-6) that these simplifying assumptions were acceptable for the problem addressed.

Use of the multiple sensor correspondence feature in conjunction with single sensor features was shown to improve performance in every measure used. The multiple sensor algorithms improved performance over the single sensor algorithms even when the single sensor algorithms were optimized. Improvements in total error rates for the range/FLIR and SD algorithms were found to be significant. These results advocate strongly for use of multiple sensors in similar problems.

The result that different sets of features gave optimum performance for different algorithms and different performance measures is most likely a consequence of ignoring statistical dependences between the features. However, the performance described in this chapter is a good estimate of how the algorithms would perform on a large amount of equivalently distributed data.

The SD algorithm merits special mention because it explicitly overcomes one limit of single sensor target detection: the ability to only detect targets segmented in the available sensor image. Thus, the SD algorithm is the recommended detection approach if detecting the most targets, in an absolute sense, is the design goal.

## VII. Conclusions and Future Directions

### 7.0 Conclusions

Use of multiple sensor information improved the performance of the target detection algorithms over the performance obtained for single sensor approaches for all comparative measures used when performance was optimized for each case. Hence, one of the fundamental hypotheses of this project was supported by the results: the hypothesis that the use of multiple sensor information can improve target detection performance. The other fundamental hypothesis of the project was also supported: the processing architecture shown in Figure (1-1) was found to provide a useful approach to extracting and processing multiple sensor information. Single and multiple sensor processes were partitioned in the architecture, allowing for the use of non-pixel registered imagery. Regions of interest were geometrically registered between the images, rather than pixels. Careful design and implementation of the multiple sensor systems was required, but this research provides concrete evidence that information only obtainable from multiple sensors can be used to improve target detection performance.

Multiple sensor information was incorporated into the target detection process through the correspondence feature. The underlying principle of the correspondence feature was that targets occupy the same space in all views of a scene, while segmented non-target regions do not tend to behave in this manner. The implementation of this principle was developed for the specific cases of FLIR and range images. However, this concept should generalize directly to other combinations of sensors. To successfully implement the correspondence feature for other combinations of sensors, the requirement is that the sensors and their associated segmentation algorithms (or region of interest selection algorithms) do not tend to provide false segmentations on similar types of scene elements.

The ability to perform multiple sensor operations by registering regions between the images is a useful departure from the more common approach of registering pixels through sensor design. Registration of regions is less physically demanding on the design of the individual sensors, allowing 'optimal' individual sensors to be built and mounted separately on a platform. The concept of optimality is used here in the sense that no design concessions need be made to the problem of sharing an aperture between the sensors. These sensors could be used to survey disjoint scenes until multiple sensor information is required, increasing coverage over an otherwise identical multiple sensor system using a single aperture. The cost of this approach is that an accurate estimate of the geometric transformation between the sensors must be maintained by the sensor positioning systems.

The processing architecture used to process multiple sensor information is generally applicable, and may find use in future systems. This architecture was demonstrated for two sensors, but is extensible to more than two sensors.

FLIR image segmentation was accomplished based on pixel brightness and heuristic operations performed on regions. The initial segmentation step, an adaptive threshold operation, used a heuristic rule to choose the threshold based on an automated inspection of the histogram of an image. This technique provided excellent performance. However, it is extensible only to FLIR images possessing approximately the same target and background brightness distributions as the data base used here.

New results in range image segmentation were obtained. Specifically, tactical targets were segmented based on the small-scale planar nature of their surfaces. Surface orientation was explicitly neglected in this technique in favor of a novel planarity test. The critical parameter in the planarity test, a threshold on the absolute error associated with fitting planes to  $3 \times 3$  regions in range images, was developed as a function of the standard deviation of the range measurements (also known as the range accuracy). The standard deviation of range measurements was shown to depend upon system perfor-

mance measures and imaging parameters. Hence, the error threshold was a function of physically significant and readily obtained measurements, a very useful property in a segmentation system. The range segmentation algorithm is extensible to other problems where small-scale planar objects are to be found in scenes which do not possess the small-scale planarity property.

Typical outputs of the segmentation systems were images which contained a large fraction of the targets present, and some regions which did not correspond to any target. The post-segmentation target detection problem was that of partitioning segmented target regions from segmented non-target regions. This problem was formulated as a two-class estimation problem, where the classes were target and non-target.

Bayesian decision theory was used to perform class estimation. The Bayesian minimum error criterion, called the Maximum a Posteriori (MAP) decision criterion, was used as the class estimation rule. The classes were assumed to be, a priori, equally likely. Class-conditioned probabilities for the features were computed by assuming conditional independence between the features and using a histogram approach to computing the conditional probabilities.

An initial set of features was evaluated for use in the class estimation system. This initial set of features was chosen based on sensor physics and an evaluation of the differences between segmented target regions and segmented non-target regions. A selection process was applied to the features to select the best three features for each type of sensor image based on the criterion of minimizing the single feature probability of classification error.

Five detection systems were developed and compared: (1) FLIR-only; (2) range-only; (3) FLIR assisted by range image information, or FLIR/range; (4) range assisted by FLIR image information, or range/FLIR; and (5) the single decision (SD) algorithm. The single sensor cases, FLIR-only and range-only, provided baseline performance for single sensor information. The multiple sensor cases, FLIR/range, range/FLIR, and SD, were

distinguished from the single sensor cases by use of multiple sensor correspondence feature information in conjunction with single sensor feature information in the class estimation process.

The FLIR/range and range/FLIR algorithms were fundamentally limited to only detecting target in the dominant sensor image. The SD algorithm overcame this limit by resolving joint spatial occupancy issues between the segmented regions in each image and making a single class estimate for each segmented region of space, regardless of whether the region was segmented in one or both sensor images. Thus, the SD algorithm was capable of detecting targets segmented in only one sensor image. The SD algorithm was shown to detect more targets than any of the other target detection approaches.

### 7.1 Future Directions

Better geometric registration will allow the correspondence feature measurement to be refined. Specifically, the various fractions of pixels in the cued regions used to declare the various values for the correspondence feature were developed, in part, as a concession to small errors in selecting the common pixel. More accurate registration would allow these fractions to be raised. One likely result is that fewer segmented non-target regions would obtain the correspondence feature values indicating joint spatial occupancy with a region in the other image, improving the ability of the correspondence feature to reject non-target regions.

Better range sensing would probably improve the performance of range segmentation and the multiple sensor processes. Dense noise spikes were a particular problem in the range imagery. The presence of these spikes hurt the range segmentation performance and impacted the settings used to measure the FLIR correspondence feature. One result was that the FLIR correspondence feature was not as 'good' as the range correspondence feature using the criterion of single feature probability of classification error. Reducing or eliminating these noise spikes would improve range segmentation performance and allow the FLIR correspondence feature measurement to be modified,

with improved performance the likely result.

The assumption of conditional independence between the features coupled with the histogram approach to exhibiting the required conditional probabilities provided good performance. Alternatives exist to exhibiting these conditional probabilities. Specifically, the maximum entropy approach (Cheeseman, 1983) offers a method of computing the required conditional probabilities which accounts for the dependences between the variables. This technique is, however, computationally expensive and may not improve performance.

No work directed at recognizing targets (for example, determining automatically that a segmented region contained a tank) was performed under this project. The problem of recognizing detected objects must, however, be addressed before truly autonomous systems, including weapons systems, are developed and fielded outside the laboratory. The present research provides one approach to a target cuer which would filter input scenes and locate promising target regions for the recognition system. Work directed at automatically recognizing detected targets remains for future investigators.

## **Appendix A: Sensor Description and Data Collection Methodology**

### **A.0 Introduction**

Appendix A provides descriptions of the sensors and methods used to collect the image data used in this project. The methods used to gather the data base of collocated FLIR and absolute range imagery are also discussed.

### **A.1 FLIR Sensor**

A modified Tank Thermal Sight (TTS) FLIR sensor was used to collect the FLIR data. The standard TTS is a variation of the Army Common Module family of FLIR sensors, and is used as the thermal imaging system on many armored vehicles. The common module family of FLIR sensors was designed with a human observer as the intended end user. Modifications were made to a standard production model TTS to make it suitable as a data collection sensor.

The TTS is a two field of view infrared sensor operating in the 8 to 12 micrometer band. The fields-of-view in an unmodified TTS are nominally 2.57 degrees (deg) vertical by 3.43 deg horizontal in the narrow field-of-view, and 7.74 deg vertical by 10.32 deg horizontal in the wide field-of-view. The pixel angular subtense is nominally square and of dimension 0.186 milliradian (mr) in the narrow field-of-view and 0.56 mr in the wide field-of-view (Dockery, 1987). The sensor has 120 detectors arranged vertically, which are scanned horizontally, with interlace, to make a 240 line image. The standard TTS has lines of 320 pixels. To reduce the effects of aliasing in the horizontal dimension, the sensor was modified to oversample each resolution element by a factor of four horizontally, with no modification of the horizontal field of view. Hence, the data collected had 1280 pixels per line (Dockery, 1987). The detector elements are capacitively coupled to preamplifiers.

The 1280 pixel per line FLIR images were too large to be viewed on any available display. To compensate, adjacent pixels in raw, 1280 pixel per line images, were averaged into a single pixel. This reduced the images to 640 pixels per line, while leaving the images oversampled by a factor of two. Disk storage requirements and run times were also reduced by a factor of two by this operation.

## **A.2 FLIR Data Collection Methods**

Many of the features used to automatically segment and classify objects in FLIR images are ultimately based on the relative brightness of collections of pixels in the image. Since the relative distributions of brightness levels may be drastically changed by the settings of the gain and brightness controls of a FLIR, the method by which the settings are chosen is quite important to a successful data collection for automatic targeting technology development.

During the data collection the gain and brightness controls on the TTS FLIR were adjusted in the following manner. A histogram of the scene to be recorded was computed on a near real time basis by 'grabbing' a frame of digitized video and performing the required computation on a resident computer. This histogram was displayed to the FLIR operator, who then adjusted the gain and brightness controls of the sensor so that minimal saturation occurred at either end of the dynamic range of the sensor, and so that the average value of the pixels was in the range 30-120. Hence, the recorded data should have very few pixels with values 0 and 255 (Dockery, 1987). Histograms computed from the raw imagery largely support this description of the sensor adjustment technique.

## **A.3 Range Sensor**

The range sensor used to collect the data base was the Tri-Service Laser Radar, developed by the Raytheon Corporation (Nettleton and Smiley, 1987). Three angular resolutions and three frame sizes were supported by the sensor. The resolution used was specified by letter (i.e., A, B, or C), and ranged from 0.05 mr to 0.2 mr. The frame size



was specified by number (i.e., 5, 6, or 7), and from 64 lines by 127 columns to 256 lines by 511 columns. Resolutions, frame sizes, and corresponding sizes of the resulting fields of view are provided in Table (A-1). The pixel data was stored in the NATO format (Bohner, 1979). Absolute range pixels were represented as 32 bit unsigned integers.

| Table (A-1). Range Sensor Imaging Parameters.<br>(Nettleton and Smiley, 1987) |                      |                             |
|---|----------------------|-----------------------------|
| Name  | Resolution<br>(mrad) | Image Size<br>(line×column) |
| 5A  | 0.20                 | 64×127                      |
| 5B  | 0.10                 | 64×127                      |
| 5C  | 0.05                 | 64×127                      |
| 6A  | 0.20                 | 128×255                     |
| 6B  | 0.10                 | 128×255                     |
| 6C  | 0.05                 | 128×255                     |
| 7A  | 0.20                 | 256×511                     |
| 7B  | 0.10                 | 256×511                     |
| 7C  | 0.05                 | 256×511                     |

#### A.4 Multisensor Data Collection Technique

The multisensor data collection took place at Ft. A.P. Hill, VA, where a variety of tactical targets and backgrounds are available for viewing. Data collections took place in the Drop Zone, a landing area for paratroops during training exercises. Tactical targets could be viewed at various ranges and aspects in a broad range of environmental conditions. Backgrounds available varied from open field to tree and shrub lines. Data collections were conducted at all times of day (Dockery, 1987; Nettleton and Smiley, 1987).

The FLIR and laser radar were mounted in separate trailers. The sensors were physically separated by approximately 5 m and were approximately 3 m above the ground. The physical mounting of the sensors allowed them to be slewed so that the scene of interest could be viewed by both sensors (Dockery, 1987).

To collect a data set, the targets were first oriented to the desired aspects. The sensors were then pointed so that a common object in the scene was roughly centered in the field of view of each sensor. Adjustment of the FLIR gain and brightness controls, as out-

lined above, was accomplished. FLIR and laser radar images of the scene were then 'grabbed' simultaneously and stored to tape. Often, two or more FLIR images were grabbed in quick succession for each range image obtained.

The targets, the viewing aspects, backgrounds, operating histories, and times of day for the data collection followed a scripted plan to accomplish defined data collection goals. Information regarding the various variables of the data collection are provided elsewhere (Nettleton, 1987).

## **Appendix B: Multiple Sensor Data Files**

### **B.0 Introduction**

Appendix B provides a list of the data sets and specific image files from the Army Center for Night Vision and Electro-Optics (CNVEO) June 1987 Multisensor Data Collection used as the data base for this project. The criterion used for selecting sets of FLIR and range images for inclusion in the data base is also discussed here.

### **B.1 Data Base**

The data used in this project consisted of 97 FLIR images and 57 range images. The number of FLIR images does not match the number of range images due to the CNVEO philosophy of generally collecting two FLIR images for each range image (see Appendix A).

Images were drawn from four data sets of the June 1987 Multiple Sensor Data Collection: DF1971, DF1671, DF1572, and DF1771. Tables (B-1) through (B-4) provide the CNVEO-assigned frame numbers (FLIR images) and file names (range images) of the corresponding image sets used in this project. These identifiers are provided in Header 2 of the NATO format tapes provided by CNVEO (Bohner, 1979).

### **B.2 Selection Criterion**

A selection process was applied to the imagery which included images which could be geometrically registered in the data base. This process involved manual review of matched sets of segmented images.

Specifically, to be included as a FLIR/range image set in the data base at least one target common to both images had to be segmented with high accuracy. This requirement allowed the selection of the center pixel in each sensor view of the common target

to be taken as originating from the same scene element, providing a basis for geometrical registration of regions between the images. This pixel was called the common pixel. Geometric registration of the images was then accomplished through angular translations from the common pixel in each image.

Table (B-1). DF1971 Multiple Sensor Data Sets

| FLIR Frame # | Range File Name |
|--------------|-----------------|
| 62311        | LD61976C09TA1   |
| 62494        | LD61976C09TA1   |
| 78264        | LD61976C11TR1   |
| 78447        | LD61976C11TR1   |
| 13567        | LD61975B11TR1   |
| 13749        | LD61975B11TR1   |
| 18939        | LD61975B10AP1   |
| 19122        | LD61975B10AP1   |
| 23108        | LD61975B09TA1   |
| 23290        | LD61975B09TA1   |
| 71348        | LD61975B10AP2   |
| 77212        | LD61975B11TR2   |
| 77395        | LD61975B11TR2   |
| 00252        | LD61976C11TR2   |
| 00434        | LD61976C11TR2   |
| 05745        | LD61976C10AP2   |
| 05927        | LD61976C10AP2   |
| 13907        | LD61976C09TA2   |
| 14089        | LD61976C09TA2   |
| 48024        | LD61976C09TA3   |
| 48207        | LD61976C09TA3   |
| 52052        | LD61976C10AP3   |
| 52234        | LD61976C10AP3   |
| 55773        | LD61976C11TR3   |
| 55956        | LD61976C11TR3   |
| 73578        | LD61975B11TR3   |
| 73760        | LD61975B11TR3   |
| 77743        | LD61975B10AP3   |
| 77926        | LD61975B10AP3   |
| 82491        | LD61975B09TA3   |
| 82674        | LD61975B09TA3   |
| 10660        | LD61975B09TA4   |
| 10843        | LD61975B09TA4   |
| 17752        | LD61975B10AP4   |
| 17934        | LD61975B10AP4   |
| 21494        | LD61975B11TR4   |
| 21677        | LD61975B11TR4   |
| 49583        | LD61976C11TR4   |
| 53288        | LD61976C10AP4   |
| 53470        | LD61976C10AP4   |

Table (B-2). DF1671 Multiple Sensor Data Sets

| FLIR Frame # | Range File Name |
|--------------|-----------------|
| 67234        | LD61677A17M1    |
| 67417        | LD61677A17M1    |
| 79473        | LD61677C17M1    |
| 96950        | LD61677A17M2    |
| 97133        | LD61677A17M2    |
| 21668        | LD61677A17M3    |
| 21850        | LD61677A17M3    |
| 69935        | LD61677C17M4    |
| 70118        | LD61677C17M4    |
| 76829        | LD61677A17M4    |
| 77012        | LD61677A17M4    |
| 23200        | LD61677C17M6    |
| 29035        | LD61677A17M6    |
| 29218        | LD61677A17M6    |
| 66473        | LD61677A17M7    |
| 66656        | LD61677A17M7    |
| 71527        | LD61677C17M7    |
| 71710        | LD61677C17M7    |
| 01448        | LD61677A17M8    |

Table (B-3). DF1572 Multiple Sensor Data Sets

| FLIR Frame # | Range File Name |
|--------------|-----------------|
| 47396        | LD61577C10M1    |
| 47578        | LD61577C10M1    |
| 53895        | LD61576B10M11   |
| 54077        | LD61576B10M12   |
| 67110        | LD61576B10M21   |
| 67293        | LD61576B10M22   |
| 71023        | LD61577C10M2    |
| 71206        | LD61577C10M2    |
| 95739        | LD61577C10M3    |
| 95921        | LD61577C10M3    |
| 01983        | LD61576B10M31   |
| 02165        | LD61576B10M32   |
| 15431        | LD61576B10M41   |
| 15614        | LD61576B10M42   |
| 19643        | LD61577C10M4    |
| 19825        | LD61577C10M4    |
| 56898        | LD61576B10M5B1  |
| 57080        | LD61576B10M5B2  |
| 67868        | LD61576B10M62   |
| 71678        | LD61577C10M6B   |
| 71861        | LD61577C10M6B   |
| 76979        | LD61577C10M6B   |
| 77161        | LD61577C10M6B   |
| 06387        | LD61577C10M7    |
| 11119        | LD61576B10M72   |
| 33781        | LD61576B10M82   |
| 37676        | LD61577C10M8    |
| 37858        | LD61577C10M8    |

| Table (B-4). DF1771 Multiple Sensor Data Sets |                 |
|---|-----------------|
| FLIR Frame #                                  | Range File Name |
| 55996   | LD61777C17M1    |
| 56180   | LD61777C17M1    |
| 29011   | LD61777C17M2    |
| 29194   | LD61777C17M2    |
| 20774   | LD61777C17M7    |
| 20956   | LD61777C17M7    |
| 53922   | LD61777C17M8    |
| 54105   | LD61777C17M8    |
| 44034   | LD61777C17M10   |
| 44216   | LD61777C17M10   |



## **Appendix C: Absolute Performance of Detection Algorithms**

### **C.0 Introduction**

Absolute performance of all the detection algorithms as a function of feature choice is provided in this appendix. Three FLIR features and three range features were used. The FLIR features were:

- 1) Complexity
- 2) Length-to-width ratio
- 3) Contrast of the means

The range features were

- 1) Length-to-width ratio
- 2) Absolute difference of the standard deviations
- 3) Complexity

These features are referred to by sensor and feature index. Thus, FLIR:1 refers to the FLIR image complexity feature. The absolute class estimation performance of each class estimation technique is reported as a function of combinations of these features.

There were 153 target opportunities and 23 non-target opportunities for the FLIR-only and FLIR/range algorithms. There were 207 target opportunities and 463 non-target opportunities for the range-only and range/FLIR algorithms. The SD algorithm had 217 target opportunities and 483 non-target opportunities.

### **C.1 Performance**

In the succeeding tables the absolute class estimation performance is reported in two categories: (1) the number of target regions correctly classified (# target regions correct); and (2) the number of non-target regions incorrectly classified (# non-target regions incorrect). In this scheme, (1) represents the detection rate and (2) is a value

required to compute the rate of false alarms per detection declaration.

| Table (C-1). Performance for FLIR:1; and Range:1. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 110                         | 1                                 |
| Range   | 176                         | 116                               |
| FLIR/Range  | 126                         | 1                                 |
| Range/FLIR  | 197                         | 30                                |
| SD  | 198                         | 32                                |

| Table (C-2). Performance for FLIR:2; and Range:2. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 120                         | 6                                 |
| Range   | 79                          | 47                                |
| FLIR/Range  | 120                         | 4                                 |
| Range/FLIR  | 197                         | 47                                |
| SD  | 194                         | 50                                |

| Table (C-3). Performance for FLIR:3; and Range:3. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 99                          | 4                                 |
| Range   | 132                         | 81                                |
| FLIR/Range  | 122                         | 3                                 |
| Range/FLIR  | 192                         | 45                                |
| SD  | 193                         | 47                                |

| Table (C-4). Performance for FLIR:1,2; and Range:1,2. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 127                         | 1                                 |
| Range   | 118                         | 92                                |
| FLIR/Range  | 129                         | 1                                 |
| Range/FLIR  | 186                         | 25                                |
| SD  | 185                         | 27                                |

| Table (C-5). Performance for FLIR:1,3; and Range:1,3. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 130                         | 1                                 |
| Range   | 176                         | 89                                |
| FLIR/Range  | 138                         | 1                                 |
| Range/FLIR  | 187                         | 29                                |
| SD  | 191                         | 31                                |

| Table (C-6). Performance for FLIR:2,3; and Range:2,3. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 127                         | 1                                 |
| Range   | 109                         | 52                                |
| FLIR/Range  | 131                         | 1                                 |
| Range/FLIR  | 177                         | 41                                |
| SD  | 180                         | 41                                |

| Table (C-7). Performance for FLIR:1,2,3; and Range:1,2,3. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 131                         | 1                                 |
| Range   | 141                         | 25                                |
| FLIR/Range  | 136                         | 1                                 |
| Range/FLIR  | 187                         | 21                                |
| SD  | 187                         | 23                                |

| Table (C-8). Performance for FLIR:1,3; and Range:1. |                             |                                   |
|---|-----------------------------|-----------------------------------|
| Algorithm   | # Target Regions<br>Correct | # Non-target Regions<br>Incorrect |
| FLIR  | 130                         | 1                                 |
| Range   | 176                         | 116                               |
| FLIR/Range  | 138                         | 1                                 |
| Range/FLIR  | 197                         | 30                                |
| SD  | 201                         | 32                                |

## **Bibliography**

- Bachman, C.G. **Laser Radar Systems and Techniques**. Dedham, MA: Artech House, Inc., 1979.
- Besl, Paul, and Ramesh Jain. "Range Image Understanding", **Proceedings of the IEEE Conference on Pattern Recognition**. 430-449. New York: IEEE Press, 1985.
- Besl, Paul J., and Ramesh C. Jain. "Segmentation Through Variable-Order Surface Fitting", **IEEE Transactions of Pattern Analysis and Machine Intelligence, PAMI-10**: 167-192 (March 1988).
- Bogler, Phillip L. "Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems", **IEEE Transactions on Systems, Man, and Cybernetics, SMC-17**: 968-977 (December 1987).
- Bohner, Manfred. "A Tape Format for Transferral of Image Data and Source Programs", **Computer Graphics and Image Processing, 11**: 185-191 (1979).
- Bullock, T.E., et al. "Sensor Fusion Applied to System Performance Under Sensor Failures", **Proceedings of the SPIE Conference on Sensor Fusion, Volume 931**. 131-135. Bellingham, WA: SPIE, 1988.
- Burden, Richard L., et al. **Numerical Analysis**. Boston: Prindle, Weber, and Schmidt, 1980.
- Cheeseman, Peter. "A Method of Computing Generalized Bayesian Probability Values For Expert Systems", **Proceedings of the Eighth International Joint Conference on Artificial Intelligence**. 198-202. 1983.
- Cheeseman, Peter. "In Defense of Probability", **Proceedings of the Ninth Conference on Artificial Intelligence**. 1002-1009. 1985.
- Comparato, Vito G. "Fusion - The Key to Tactical Mission Success", **Proceedings of the SPIE Conference on Sensor Fusion, Volume 931**. 2-7. Bellingham, WA: SPIE, 1988.
- Devijver, P.A. and J. Kittler. **Pattern Recognition - A Statistical Approach**. Englewood Cliffs, NJ: Prentice/Hall International, 1982.
- Dockery, Robert, Project Engineer. Personal interview. Army CECOM Center for Night Vision and Electro-Optics, Ft. Belvoir, VA, 30 October 1987.

Duane, Greg. "Pixel-Level Sensor Fusion for Improved Object Recognition", **Proceedings of the SPIE Conference on Sensor Fusion**. 180-185. Bellingham, WA: SPIE, 1988.

Duda, Richard O. et al. "Subjective Bayesian methods for rule-based inference systems". **Proceedings of 1976 American Federation of Information Processing Societies (AFIPS) National Computer Conference**. 1075-1082. Montvale, NJ: AFIPS, 1976.

Duda, Richard O. et al. "Use of Range and Reflectance Data to Find Planar Surface Regions", **IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1**: 259-271 (July 1979a).

Duda, R.O. et al. **Computer-Based Consultant for Mineral Exploration: Final Report**. Contract AER77-04499. Menlo Park, CA. SRI International, September, 1979b (NSF/RA-790228).

Due, Christopher T., and Lauren M. Peterson. **Optical-Mechanical, Active/Passive Imaging Systems - Volume I**, Contracts N00014-77-C-0125, N00014-80-C-0510, N00014-81-C-0425. Ann Arbor Michigan: Environmental Institute of Michigan, Infrared Information and Analysis (IRIA) Center, May 1982. Available only, at no cost, from the IRIA Center, P.O. Box 8618, Ann Arbor, Michigan.

Foley, Donald H. "Considerations of Sample and Feature Size", **IEEE Transactions on Information Theory, IT-18**: 618-626 (September 1972).

Fukunaga, Keinosuke. **Introduction to Statistical Pattern Recognition**. New York: Academic Press, 1972.

Garvey, Thomas D., and John D. Lowrance. **Machine-Intelligence-Based ESM System, Interim Technical Report**. Contract F33615-80-C-1110. Menlo Park, CA: SRI International, December 1981 (SRI Project 1655).

Gonzalez, Rafael C., and Paul Wintz. **Digital Image Processing**. Reading, MA: Addison-Wesley Publishing Co., 1987.

Harmon, William, Project Engineer. Personal Interviews. Air Force Wright Aeronautical Laboratories, AFWAL/AARI, Wright-Patterson AFB, OH, November 1987 - December 1988.

Haskins, Thomas G. "Sensor Cueing Performance Analysis", **Proceedings of the NAE-CON Conference**. 262-265. New York: IEEE press, 1985.

- Hoffman, Richard, and Anil K. Jain. "Segmentation and Classification of Range Images", **IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9**: 608-620 (September 1987).
- Keeping, E.S. **Statistical Inference**. Princeton, NJ: D. Van Nostrand Company, Inc., 1962.
- Kreigman, David J., et al. "Mobile Robot Planning with Integrated Sensor Input", **Proceedings of the 26th IEEE Conference on Decision and Control**. New York: IEEE Press, 1982.
- Lewis, P.M. "The Characteristic Selection Problem in Recognition Systems", **IRE Transactions on Information Theory, 8**: 171-178 (1962).
- Lloyd, J.M. **Thermal Imaging Systems**. New York: Plenum Press, 1975.
- Lowrance, John D., and Thomas D. Garvey. **Evidential Reasoning: An Implementation for Multisensor Integration**. SRI Technical Note 307. Menlo Park, CA: SRI International, December 1983.
- Magee, M.J., and J.K. Aggarwal, "Using Multisensory Images to Derive the Structure of Three-Dimensional Objects: A Review", **Computer Vision, Graphics, and Image Processing, 32**: 145-157 (November 1985).
- Melsa, James L., and David L. Cohn. **Decision and Estimation Theory**. New York: McGraw-Hill Book Co., 1978.
- Milgram, D.L., and C.M. Bjorklund. "Range Image Processing: Planar Surface Extraction", **Proceedings of the Fifth International Conference on Pattern Recognition**. 912-915. New York: IEEE Press, 1980.
- Mitiche, Amar, and J.K. Aggarwal. "Multiple Sensor Integration/Fusion Through Image Processing: A Review", **Optical Engineering, 25**: 380-386 (March 1986).
- Nettleton, John E., and Alan Smiley. **Multisensor Field Test Laser Radar Log, Ft. A.P. Hill, June 1987**. Army CECOM Center for Night Vision and Electro-Optics, Laser Division, Ft. Belvoir, VA, 1987.
- Nettleton, John E., Electrical Engineer, Personal Correspondence. Army Center for Night Vision and Electro-Optics, Laser Division, Fort Belvoir, VA, 21 March 1989.
- Papoulis, Athanasios. **Probability, Random Variables, and Stochastic Processes**.

New York: McGraw-Hill Book Co., 1965.

Roggemann, Michael C., et al. "Multisensor Information Fusion for Target Detection and Classification", **Proceedings of the SPIE Conference on Sensor Fusion, Volume 931**. 8-11. Bellingham, WA: SPIE, 1988.

Rosenfeld, Azriel, and Avinash C. Kak. **Digital Picture Processing, Volume 2**. Orlando, FL: Academic Press, Inc., 1982.

Ruck, Dennis W., et al. "Multisensor Target Detection and Classification", **Proceedings of the SPIE Conference on Sensor Fusion, Volume 931**. 14-21. Bellingham, WA: SPIE, 1988.

Teague, Michael Reed. "Image Analysis via the General Theory of Moments", **Journal of the Optical Society of America**, 70: 920-930 (August 1980).



## VITA

Captain Michael C. Roggemann was born on 1 September 1960 in Hampton, Iowa. He graduated from high school in Charles City, Iowa in 1978, and attended Iowa State University, receiving the degree of Bachelor of Science in Electrical Engineering in May 1982. Upon graduation he received a commission in the USAF through the ROTC program. Captain Roggemann entered active duty in May 1982 as a graduate student in the electro-optics program at AFIT, receiving the degree of Master of Science in Electrical Engineering in December 1983. From December 1983 to June 1986 he was assigned to the Air Force Wright Aeronautical Laboratories, Avionics Laboratory, Electro-Optics Branch. In June 1986 Captain Roggemann entered AFIT again to begin work toward the degree of Doctor of Philosophy.

END

6-89

DTic